

CLASSIFICATION:
PRIOR APPLICATION DATA: US 60/042,427
FILING DATE: 27-MAR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/036,205
FILING DATE: 27-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/035,920
FILING DATE: 22-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hibler, David W.
REGISTRATION NUMBER: 41,071
REFERENCE/DOCKET NUMBER: UTSD:537
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 13865 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-009-656-11

QY	1669	taaatataaattggtgccaggccgggtgqgtgtgctcatgctctgtaatccccagcat	1728
Db	8368	TTAAGAAAATATAAGAAAGAGGCTGGGCACTGGTGGCTGTAATCTCAGAAC	8427
QY	1729	tttgagaggtgaggggtggatcacctgaggtcagga-ttcaaaaccagcctggccaa	1787
Db	8428	TTTGGGATGCTAAGGCGGGTGGATCACTTGAGGTGAGGAGTCAAGACCAAGCCTGGCCAA	8487
QY	1788	catagtgaacccccctctctactaaaaatgcaaaaattagtcgqgctgggtggcacactc	1847
Db	8488	CATGGTGAACCCCATCTTACTAAAAATCAAAAATTGAGCCAGCGGTTGTGGGTGC	8547
QY	1848	ctgtaatccccagcttaacttaggaggtgaggtgggaaatactctggacctccaaaggtgga	1907
Db	8548	CTGTAAATCCACAGCTACTTTGGGAGGCTGAGGAGAGGAATCACTTCGACCCAGGTGGAGGA	8607
QY	1908	ggttgagtagtgagtgaaatcatgcccactgacccctagcttggtggcgcaaaacactct	1967
Db	8608	GGTTGCAGTGAACGTAGATTGCGCCACTGCGCACTCGGGCTGGTGACAGCGAGACTCT	8667
QY	1968	atcaaaaaataataataattgtttcaagtcctgccaaaaaataaaaaa	2027
Db	8668	GTCIAAAAGAAAGAGGGAAAGAAAGAAAAGAAAAGAAAGAAAAGAAAGAGGAG	8727
QY	2028	aaaaaaaaaaaaaaaaa	2044
Db	8728	GAAGAGAAAGAAATATA	8744

RESULT 5
 US-09-009-913-1/c
 ; Sequence 1, Application US/09009913
 ; Patent No. 6087485
 ; GENERAL INFORMATION:
 ; APPLICANT: Axys Pharmaceuticals, Inc.
 ; TITLE OF INVENTION: Asthma Related Genes
 ; NUMBER OF SEQUENCES: 339
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bozicevic & Reed, LLP
 ; STREET: 285 Hamilton Ave, Suite 200
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA

STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,901
FILING DATE: Filed herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PP-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2713 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-916-901-6

Query Match	9.9%;	Score 202.8;	DB 2;	Length 2713;
Best Local Similarity	80.6%;	Pred. No. 7.1e-31;		

	Matches	274; Conservative	0; Mismatches	62; Indels	4; Gaps	3;
QY	1664	actgtcaaaataaat	tgtgcacgcccgggtgcggtggctatgctgaattccc		1723	
Db	1086	ACTGTTTCATGTAA	AAGAATGGTTCGGCCAGCGCGTGCTCATGCTGTAATGCC		1145	
QY	1724	agcatttggaggctg	gagcggttgcatcacctgaggtcagga-ttcaaaaccagcctg		1782	
Db	1146	AGCACTTTGGAGC	CGCCAGCGSGTGCATCACCTGAGGTCAGGATTCGACACCAGGCTG		1205	
QY	1783	gccacaatatgaa	cccccc-ctctctaataaaatgc--aaaaattagtgcgggtggtg		1839	
Db	1206	GCCAACAATGTTGA	ACCCCCTGCTCTACTAAAAATACAAAATAATAGTCGGCGCGGTG		1265	
QY	1840	gcacacctctgtat	ccocagctaacttaggaggtgaggtgggaaaaaacactttggaotcca		1899	
Db	1266	GCGCANGCCTGTAT	CCTCCAGCTACTCAGGAGGCTGAGGCGAGGAGAACACTTGAACCCGG		1325	
QY	1900	aaggtggaggttgc	agttaagctgaaatcatgcacctgcacctagcttgggtggcaagc		1959	
Db	1326	GAGGCAGAGGTTGC	ATGAGCCAGATCATGCCATTTCACCTCAGGCTGGGTGACAGAGC		1385	
QY	1960	aaacrtctatcaaa	aaaaataataataattgttcaaaa		1999	
Db	1386	GAGACTCCATCTCA	AAAAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA		1425	

```

RESULT      7
US-09-154-602-6
; Sequence 6, Application US/09154602
; Patent No. 630472
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,602
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,901
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0367 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2713 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-154-602-6

Query Match 9.9%; Score 202.8; DB 4; Length 2713;
Best Local Similarity 80.6%; Pred. No. 7.1e-31;
Matches 274; Concentration 0.

Accession	274;	Conservative	0;	Mismatches	62;	Indels	4;	Gaps
QY 1604	actgctaaataataatgtgccaggccgggtgcgggtcattcgctgtaatccc							1723
DG 1086	ACTGTTCATTGAAAAGAAATGGTCTCCGCCAGCGCGGTGGTCATGCCTGTAAATCCC							1145
QY 1724	agcaattttgagagctgagggcgttggaaccctgagtacagga-ttcaaacaacagacctg							1782
DB 1146	AGCACTTTGGAGCGCCGAGCGGGTGGAATCACCTGAGTTCAGGAGTTCGAGACGACCTG							1205
QY 1783	gccaacatagtgaaacccc-ctctctactaaaaatgc--aaaattagtcgggcggtggtg							1839
DB 1206	GCCAACAATGGTGAACCCCTGCTCTACTAAAAATACAAAAAATAATGCTGGCGCGGTG							1265
QY 1840	gcacaactcctgtaatccacagctacttaggagctgaggtggggaaatacaacttggactcca							1899
DB 1266	CGCATGCCCTGTAAFPCCAGCTACTCAGGAGGCTGAGGAGAGAATCACTTGAACCCGG							1325
QY 1900	aagtgtaggtgtgcagtaagctaaatacatgccactcacccctagcttggtggcaaacg							1959
DB 1326	GAGSACAGAGGTTCAGTTCAGCCGAGATCATGCCCTTCATCTCAGCCTGGGTGACAGAG							1385
QY 1960	aaaactctatcaaaaaaaaaataataaatattgttcaaaa							1999
DB 1386	GAGACCACTCTCAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA							1425

RESULT 8
US-08-395-800A-7
; Sequence 7, Application US/08395800A
; Patent No. 5807732
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B
; APPLICANT: LENNON, GREGORY
; APPLICANT: ROQUIER, SYLVIE
; APPLICANT: ROQUIER, DOMINIQUE
; APPLICANT: KELLY, ROBERT J
; TITLE OF INVENTION: GDP-L-FITCOPSE
REVA-D-CATACOMOSTE

Query Match	9.6%	Score 196.8;	DB 1;	Length 2354;
Best Local Similarity	52.4%;	Pred. No. 9.9e-30;		
Matches 575; Conservative	0;	Mismatches 487;	Indels 36;	Gaps 5;
QY	269	cagaccacctgctctgtcttctatcatctccgcgcaggccctatccccagcggc	328	
Dd	475	CTGAGGACGGGTGCACAGCTACCTTCAGTTGCTGGGAGCTGGGCATACGACC	534	

...

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:35:34 ; Search time 4771.45 Seconds
(without alignments)
4055.621 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggcgcacaggttctc.....gtgggcacgcctgactga 1173

Scoring table: IDENTITY_NDC
Gap 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Result No.	Score	Match	Length	DB	ID	Description
1	1173	100.0	1462	9	AF176839	AF176839 Homo sapi
2	1173	100.0	2170	9	AF176838	AF176838 Homo sapi
3	1173	100.0	3278	9	AF246718	AF246718 Homo sapi
4	1173	100.0	71503	9	AF219991	AF219991 Homo sapi
5	1173	100.0	159072	2	AC026419	AC026419 Homo sapi
6	1173	100.0	181566	2	AC025387	AC025387 Homo sapi
7	1053.8	89.8	179065	2	AC009163	AC009163 Homo sapi
8	924.6	78.8	2544	9	AF219990	AF219990 Homo sapi
9	924.6	78.8	3786	9	AF280086	AF280086 Homo sapi
10	924.6	78.8	159072	2	AC026419	AC026419 Homo sapi
11	924.6	78.8	208185	2	AC009105	AC009105 Homo sapi
12	745.2	63.5	1740	10	AF176841	AF176841 Mus muscu
13	745.2	63.5	1989	10	AF176840	AF176840 Mus muscu
14	400	34.1	2032	9	AF131235	AF131235 Homo sapi
15	398.4	34.0	1333	9	AF149783	AF149783 Homo sapi
16	398.4	34.0	1992	9	AF280088	AF280088 Homo sapi
17	398.4	34.0	2011	9	AK026635	AK026635 Homo sapi
18	398.4	34.0	183133	2	AC010547	AC010547 Homo sapi
19	358.8	30.6	2201	10	AF109155	AF109155 Mus muscu
20	354	30.2	1926	10	AF131236	AF131236 Mus muscu
21	351.6	30.0	179065	2	AC009163	AC009163 Homo sapi
22	235.2	20.1	2156	6	AR071396	AR071396 Sequence
23	235.2	20.1	2156	6	E14937	E14937 Human mRNA
24	235.2	20.1	2731	9	AB012192	AB012192 Homo sapi
25	235.2	20.1	6961	9	AB017915	AB017915 Homo sapi
26	235.2	20.1	182677	2	AC073370	AC073370 Homo sapi
27	235.2	20.1	196465	9	AC022392	AC022392 Homo sapi
28	196.8	16.8	2354	5	CHKG6ST	CHKG6ST
29	196.8	16.8	2354	6	AR050577	AR050577 Sequence
30	196.8	16.8	2354	6	E12369	E12369 DNA encodin
31	196.2	16.7	137499	2	ACU15931	ACU15931 Homo sapi
32	196.2	16.7	165762	2	AP001582	AP001582 Homo sapi
33	196.2	16.7	174187	2	AP002792	AP002792 Homo sapi
34	196.2	16.7	215647	2	AC068591	AC068591 Homo sapi
35	190.6	16.2	1425	10	AF178689	AF178689 Rattus no
36	181	15.4	1654	10	AB008937	AB008937 Mus muscu
37	153.8	13.1	2227	9	AF083066	AF083066 Homo sapi
38	153.8	13.1	3029	9	AK023268	AK023268 Homo sapi
39	153.8	13.1	3535	9	AB021124	AB021124 Homo sapi
40	153.8	13.1	4862	9	AB021125	AB021125 Homo sapi
41	153.8	13.1	169753	2	AC055737	AC055737 Homo sapi
42	150.8	12.9	182575	2	AC018450	AC018450 Homo sapi
43	148.8	12.7	2393	9	AB014679	AB014679 Homo sapi
44	148.8	12.7	2393	9	AB014680	AB014680 Homo sapi
45	148.8	12.7	2409	6	E28527	E28527 Polypeptide

ALIGNMENTS

RESULT 1

AF176839

LOCUS

DEFINITION

Homo sapiens intestine N-acetylglucosamine 6-o-sulfotransferase (I-GlcNAc-6-ST) gene, complete cds.

ACCESSION

AF176839

VERSION

AF176839.1

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1462)

Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.

Cloning and characterization of a mammalian

N-acetylglucosamine-6-sulfotransferase that is highly restricted to

intestinal tissue

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

99423499

REFERENCE

2 (bases 1 to 1462)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
/tissue_type="intestine"
1. .2170
/gene="I-GLNAC-6-ST"
344. .1516
/gene="I-GLNAC-6-ST"
/notes="possibly involved in biosynthesis of mucosal
ligands for L-selectin"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56000.1"
/db_xref="GI:5917706"
/translation="MMLPRFSKTVVLLIAOTTCILLFTISRPSSPAGBDRVHV
LVLSWRSGSLGOLFSDPVDYVILMEPAHWVITLSQSAATLHMVNDLMKSLFL
CMQVDFADYMPQSRNLSEAFENWATSEALCSFAPRGTISKQDVCTLCIRQPF
LARECRSYSHVILKEVRFNQLVPLSLPALNRLVHLVRDPAVLRSRAAGPI
LARDNGIVLGTNGKWEADPHRLRKEVCRSHVRIAEATLKPPPLRGRRLVRFD
LAREPIAEIRALYAFGLTLPLEAWIHNTHGSGIGRPIEAFHPSNRNRYNSQAW
RHALPFTKILRYQVEVCAGALQLLGRPVYSADQQRDLTLDLVLPRDFHSWASPD"
BASE COUNT 449 a 582 c 612 g 427 t
ORIGIN

Query Match 100.0%; Score 1173; DB 9; Length 2170;
Best Local Similarity 100.0%; Pred. No. 3.8e-161;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtgctcagcgttctccagcaagacagtgacagtgctctctctctctctcagcacc 60
Db 344 ATGTGCTGCCACGGTCTCCAGCAGACAGTACAGTGTCTCTCTCTCTCTCTCTCTCT 403

QY 61 tgcctctgtcttctatctatctcccgccagcgccctctctctctctctctctctctct 120
Db 404 TGCCTCTGTCTTCTATCATCTCCGGCCAGCGCCCTCATCCCGCCGCGCGCGAGGAT 463

QY 121 cgtgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
Db 464 CTGTGTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 523

QY 181 ttcaagccagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
Db 524 TTCAGCAAGCAGCCCGAGCTCTTACCTGATGAGCGCGCGGTGTCATGTGAGCAGCACC 583

QY 241 cgtgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 300
Db 584 CTGTGTGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360

QY 301 ttttgtgacatgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
Db 644 TTTTGTGGACATGGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 703

QY 361 ttttcaactgggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
Db 704 TTTTCAACTGGGACAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 763

QY 421 cagggcaccatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
Db 764 CGAGGCACCATCAGCAAGCAGGAGGTATGCAAGACATGTGTGTGTGTGTGTGTGTGTGTGT 823

QY 481 ctggcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
Db 824 CTGGCCGGGAGGCGCTGCGCGCTCTACAGCACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 883

QY 541 aactgcaggtgtctaccgcgtactcagcagcagcagcagcagcagcagcagcagcagcagc 600
Db 884 AACCTGCAGGTGTCTACCGCGTGTCTACGAGCAGCCCGCGGCTCAACCTGTGTGTGTGTGT 943

QY 601 ctggtgcgacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
Db 944 CTGGTGGCGACCCCGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1003

```

```

QY 661 cgcgacaacggcatcgtgctggggcaccacagcgaagtgggtgagggcaccacctcaactg 720
Db 1004 CCGGACAACGGCATCGTGTGGGACCAACGCAAGTGGTGGAGGCGCACCTCACCTG 1063

QY 721 gcctgattcggaggtgtgcgcagccacgtgcgcacgtgcgcacgtgcgcacgtgcgcac 780
Db 1064 CGCCTGATTTCGCGAGGTGTGCGCAGCGACGCTGCGCATCGCGAGGCGCGCACATCAAG 1123

QY 781 cgcgcacacctctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 1124 CGGCACACCTTCTTCCGCGCGCGCTTACCGCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1183

QY 841 ccgtggcagagatccgcgcacacttaacgcttcaacgacgtgaacctcaacgacacagctc 900
Db 1184 CCGCTGGCAGAGATCCGCGCATCTACGCCCTTACCGGCCCTGACCCCTCACGCCACAGCTC 1243

QY 901 gagccttgatcccaacaacatcaccacacgggtgcgggagtcggaagccaatcgagccttc 960
Db 1244 GAGCCTTGGATCCCAACATCACCACCGGTGCGGATCGCAAGCCAATCGAGGCTTC 1303

QY 961 catacttctctaggaatcgcaacgctctccagcgcctgagcgcacgogtgccttc 1020
Db 1304 CATACTCTCTAGGAATGCGCAACGTCTCCAGGCTTCCAGGCTTGGCGCCAGCTTGCCTTC 1363

QY 1021 actaagatcctgcgcgtgcagggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
Db 1364 ACTAAGATCCTGCGGTGCGAGGAGTGTGCGCGCGCGCTTCCAGGCTTGGCGCTTACCGG 1423

QY 1081 cctgttactctcggaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1140
Db 1424 CCGTGTACTCTCGGACCAACAGCAGCGTGAACCTCACCTGATCTGTGTGTGTGTGTGTGTGTGT 1483

QY 1141 ccagaccacttcagctggcgcacgtgcgcacgtgcgcacgtgcgcacgtgcgcacgtgcgcac 1173
Db 1484 CCAGACCACTTCAGCTGGGCATCGCCTGACTGA 1516

RESULT 3
AF246718 3278 bp mRNA PRI 31-OCT-2000
LOCUS Homo sapiens intestinal GLNAC-6-sulfotransferase (CHST5) mRNA,
DEFINITION complete cds, alternatively spliced.
ACCESSION AF246718
VERSION AF246718.1 GI:11055254
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3278)
Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K.,
Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A.
and Fukuda,M.N.
Macular corneal dystrophy type I and type II are caused by distinct
mutations in a new sulfotransferase gene
Nat. Genet. 26 (2), 237-241 (2000)
20472330
PUBMED 11017086
REFERENCE 2 (bases 1 to 3278)
AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K.,
Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
Tanigami,A., Thonar,E.J.-M.A., Shimomura,Y., Kinoshita,S.,
Tanigami,A. and Fukuda,M.N.
Direct Submission
Submitted (17-MAR-2000) Glycobiology Program, The Burnham
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
1. .3278
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"

```

	/map="l6q22; between D16S512 and D16S515"
gene	1..3278
5'UTR	/gene="CHST5" 1..1396 /gene="CHST5" /note="alternatively spliced" 1397..2632 /gene="CHST5" /note="I-GlcNAc6ST; alternatively spliced" /codon_start=1 /product="intestinal GlcNAc-6-sulfotransferase" /protein_id="AAG28023.1" /db_xref="GI:11055255" /translation="MGMARVPVVAHSTRPAPARMWLPFRFSKIVTVLLLAQTCLL LFILSRPPSPAGGEDRVHIVLSWRSGSFLGOLFQHPDVFYLMPEPAWEVHTTL SQGSAAITLMAVRDLMKISIFLCMDREVDAYMQPSRLNLSAFKNATSRALCSPACSAF PRGTISKQDYCKTICTQRPPLSLADEACRSYSHVVLAKEVRFNLVLVYPLLSPALNLR IVHLVRDPFVLGRSRAAGPIILDARDINGVLVTNGKWADPHLRGLREVCNRSHVRIAR AATLVPPFLKRVLYRFEDLAREPLAEIRALYAFTGLTLPQLPEAWHNTHHGSGI GKPTAEPTSSRNARNYSOARHALPETKLIRVQEVCAGALOLLGYRPVYSADQQORDL TLDIVLRGPDHFWSAPD"
CDS	BASE COUNT 686 a 988 c 866 g 738 t
	RDIGIN
	Query Match 100.0%; Score 1173; DB 9; Length 3278; Best Local Similarity 100.0%; Pred. No. 3.le-161; Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
y	1 atgtgctgccagggttctccagcaagacagtgcagtgctctctctggcacagaccac 60
b	1460 ATGTGGCTGCCAGGTTCTCCAGAAGACAGTGCAGTGTCCTCTCTGGCACAGACC 1519
y	61 tgctctgtcttcatcatcttcocggccaggcgccatccccagcggcgagagat 120
b	1520 TGCTCTGTGCTCTCATCATCTCCCGCCAGGGCCCTCATCCCCAGCGCGGCGAGGAT 1579
y	121 cgtgtcacagtgtgtgtctctgtgagcgtcgggctcatctcttgtggcagctc 180
b	1580 CGTGTGCACGTGTGGTGGTGTCTCTGTGGCGCTCGGGTCTCATCTTCTTGGCCAGCTC 1639
y	181 ttcaagccagcaccccacgctttctacctgatggagcccgctgggaatgtgtggaccac 240
b	1640 TTCACCCAGCACCCCAGAGCTTCTACTCATGGAGCCCGCTGGCATGTGTGGACCAC 1699
y	241 ctgtcgcaggcgagcgcggcaacgctgcacatggccgctgcgcagactgatgcctctac 300
b	1700 CTGTGCGAGGGCAGCGCGGGCAACGCTGCATATGGCCGTGCGGACCTGATGGCTCTATC 1759
y	301 tttittgtgcacatggagcgtgtttgatgctcatgcacagagccgaacacctgtccgc 360
b	1760 TTTTGTGCGACATGGACGCTTTTGAATGCTTACATGCCACAGACCGAAACCTGTCCGCC 1819
y	361 ttttttaactgggcaacgagcgcgcgtgtgtctgcgcgcccgccctgcagcgcttccc 420
b	1820 TTTTTCACATGGGCACAGAGCGCGCGCTGTGTCTCGCCCGCGCTGCAGCGCTTTTCCC 1879
y	421 cgaggcaccta tcagaagcaggacggtatgaagaactgtgcacgagcagccattcagc 480
b	1880 CAGGACCATCAGCAACAGAGAGGTATGCAAGACATGTGCACGCGGCGAGCCATTTCAGC 1939
y	481 ctggcccgaggagcctgcgcctctctcagccacgttgtgtctaaaggaggtgcgtctctc 540
b	1940 CTGGCCCGGAGGGCTGCGCGTCTCCCTACAGCCACGTGGTGTCTCAAGAGGTCGCTTCTC 1999
y	541 aaacctgcaggtgtctatacccgctgtctcagcagaccoccgctccaacctgcgcatcgtcac 600
b	2000 AACCTGCAGGTGCTTACTACCGCTGTCTCAGCACCCCGCGCTTCACACTGCGCATCGTGCAC 2059
y	601 ctggttcgcgaccccgcgcgctgtcgtcgtccggaggcgcgcggcccgatactggca 660
b	2060 CTGGTTCGCGACCCCGCGGCGCGTGTCTGCGCTCCCGGGAGGCGCGGGCCCGCAACTGGCA 2119

```

/rpt_type-dispersed
/rpt_family="Alu"
complement(462..771)
repeat_region

/rpt_type-dispersed
/rpt_family="Alu"
772..945
repeat_region

/rpt_family="L2"
/rpt_type-dispersed
1169..1338
repeat_region

/rpt_family="MIR"
/rpt_type-dispersed
complement(1553..1804)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
2616..2916
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
3309..3532
misc_feature

/rpt_type-dispersed
3865..4028
misc_feature

/rpt_family="Alu"
/rpt_type-dispersed
4099..4406
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
4793..5250
repeat_region

/rpt_family="L2"
/rpt_type-dispersed
complement(5251..5329)
repeat_region

/rpt_family="MIR"
/rpt_type-dispersed
complement(5768..6080)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
complement(6081..6265)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
6111..6567
misc_feature

/rpt_type-dispersed
8035..8350
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
10545..11049
misc_feature

/rpt_type-dispersed
11043..11195
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
11200..11222
repeat_region

/rpt_family="Simple_repeat"
/rpt_type-tandem
11570..11871
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
complement(12065..12198)
repeat_region

/rpt_family="L1"
/rpt_type-dispersed
complement(12205..12493)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
complement(12813..13110)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
join(14185..14228,16169..16263,17712..20821)
mRNA

/gene="CHST5"
/product="intestinal N-acetylglucosamine
6-O-sulfotransferase"
14185..20821
gene

14681..14713
repeat_region

/rpt_family="Simple_repeat"
/rpt_type-tandem
complement(14715..14990)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
complement(14995..15288)
repeat_region

```

```

/rpt_family="Alu"
/rpt_type-dispersed
complement(15292..15589)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
15568..15718
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
15977..16098
repeat_region

/rpt_family="MIR"
/rpt_type-dispersed
complement(17298..17433)
repeat_region

/rpt_family="MIR"
/rpt_type-dispersed
complement(18001..18310)
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
complement(18420..18591)
repeat_region

/rpt_family="MIR"
/rpt_type-dispersed
18732..18767
repeat_region

/rpt_family="Simple_repeat"
/rpt_type-tandem
18967..20202
CDS

/gene="CHST5"
/note="I-GlcNAc6ST; carbohydrate sulfotransferase 5;
expressed in the small intestine and colon"
/codon_start=1
/product="intestinal
N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="AAC26326.1"
/db_xref="GI:11023148"
/translation="MGMRARVPKVAHSTRPPARMMWLPFSSKTVTVLLAQTTCLL
LFIIISRPGPSFAGDEKRVHVLVSSWSGSSEIGQLFSQHPDVFYLMGPAMHVWFTL
SQGSAATLHMAYRDMIRSFIFLDMDFDAYMPQSRNLSAFEFNWATRALCSPACSAF
PRGTISKODVCKTILCTQPFSIAREACRSYSHVNLKEVFNQLVLYPLISDPALNR
IVHLVRDRAVLRSREAGPTIARDNGIVLGIKNGKVAADPHLRILREVCRRHRTAE
AATLKPPFLGRYLVRFEDLAREPLAEIRALVAFGLTLPQLKAWIHNITHGSGI
CKPIEAFHTSSRNARNVSOAWEHALPTKILRVQEVCAQALQLLGYRPPYSADQQRDL
TLDLWIPRGDPHFWSWSPD"
20504..20806
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
20869..20961
repeat_region

/rpt_family="L1"
/rpt_type-dispersed
20962..21200
repeat_region

/rpt_family="MER4-group"
/rpt_type-dispersed
21325..21429
repeat_region

/rpt_family="MER4-group"
/rpt_type-dispersed
21430..21722
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
21725..22007
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
22008..22040
repeat_region

/rpt_family="Simple_repeat"
/rpt_type-tandem
22041..22313
repeat_region

/rpt_family="MER4-group"
/rpt_type-dispersed
22325..22593
repeat_region

/rpt_family="Alu"
/rpt_type-dispersed
22594..22949
repeat_region

/rpt_family="MER4-group"
/rpt_type-dispersed
23116..23221
repeat_region

/rpt_family="Simple_repeat"
/rpt_type-tandem

```

[illegible]

Db 19890 CGCGAACGGCATGCTGCGCACCAACGGCACTGGTGGAGCGCCGACCTCACCTG 19749

RESULT	9	
AF280086	3786 bp	mrna
LOCUS		PRI 20-FEB-2001
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase GST-4beta mRNA, complete cds.	
ACCESSION	AF280086.1	GI:12060803
VERSION		
KEYWORDS		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 3786) Hemerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and Rosen,S.D.	
TITLE	Chromosomal localization and genomic organization for the galactose/N-acetylgalactosamine/N-acetylglucosamine 6-O-sulfotransferase gene family	
JOURNAL	Glycobiology 11 (1), 75-87 (2001)	
MEDLINE	21096027	
REFERENCE	2 (bases 1 to 3786) Hemerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and Rosen,S.D.	
AUTHORS	Direct Submission	
TITLE	Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA	
JOURNAL	Location/Qualifiers	
FEATURES	1..3786	
source	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="18" /map="16q23.1" 1..3786	
gene	/gene="GST4beta"	
	/note="clusters with the intestinal N-acetylglucosamine 6-O-sulfotransferase (GST4alpha) gene in GenBank Accession Numbers AF176838 and AF176839"	
CDS	189..1376 /gene="GST4beta" /codon_start=1 /product="N-acetylglucosamine 6-O-sulfotransferase GST-4beta" /protein_id="AAG48244.1" /db_xref="GI:12060804"	
	translation="FWLPRVSTAVTALLAQTLLFLVSRQSPSPAGGEARVHVLVLSRSGSFFVGFOLFNDHPDVFYILMPFAHWITLISQSAATLHMVNDLVRSVFLCDMDVFDAYLPMRNLSDFONAVSFALCSPACSPFRGALISSEYVCKPLCARQSEFLARACRSYSHVVLKVPVFNLQVLPILSDPALMRIVHLVRPRAVLKSRPQTAKALARDNGVLVGTGHWDAEDPLKRVREVCVRVIAEATILKPPFFLGRYRIVRPFDLAREFLAIFRYALVGLSLATPDLFAWHINITHSGPGARREAFKTSRRALNVQSAWRHALPFAKIRIVYFELCAGALQLLGVPRVYSEQRNLALDVLPRGLNGFTWASSTASHFRN"	
	polyA_signal 3106..3111 /gene="GST4beta"	
	/note="alternative signal"	
	polyA_signal 3734..3739 /gene="GST4beta"	
BASE COUNT	848 a 1030 c 1104 g 804 t	
ORIGIN		
Query Match	78.8%; Score 924.6; DB 9; Length 3786;	
Best Local Similarity	87.8%; Pred. No. 2.7e-125;	
Matches 1021; Conservative	0; Mismatches 139; Indels 3; Gaps 1;	
QY	1 atgtgcctgcacaggttctcacacagacagtgacagtcctctctctctggtcacagaccac 60	
Db	189 ATGTGCTGCCGCGGCTCTCCACGACGAGGACCGGCTCTCTCTCTGCGCGACG--ACC 245	
QY	61 tgctcctgcgtctttcatctcatctccggcgacagggcctcaccgccagccggcgagat 120	

D	b	246	TTCCTCCGTCCTCTTTTCGTGGTTCGCCGCAGAGGCCCTCGTCCCACGACGAGCGCGAGGCG	305
Q	y	121	cgtgtacacgctgctgtgtcgtccctcgtggcgctcggtcctcatcttcttggtgcagactc	180
D	b	306	CSCGTGCATCTGCTGTGTGCTCCTGTGGTGCTCGGGCTCGTTCGTTGGGCCAACTC	365
Q	y	181	ttagcaccaaccacgaagtcttatcctgatgaagcccagtcgtatgctgtgacaacc	240
D	b	366	TTCAACCAGCACCCGACGCTCTTACCTAATAGAGGCCGCGTGGACGCTGTGACCAACC	425
Q	y	241	ctgtcgaggggcagcgcgcgaacgcctgcacatggccgtgagcgacctgatgcctctatc	300
D	b	426	CTGTCCAGGGACGCGCGCAACGCTGCACATGGCTGTGCGCAGACTGGTGGCTCCGTC	485
Q	y	301	tttttgtgcacatgaagctgtttgatccctaacatgacacagaaccgaacactgtccgc	360
D	b	486	TTCCCTGTGCGACATGACGCTTTTGATGCCATCTATCGCTTTGGCGCCGCAACCTGTCCGAC	545
Q	y	361	tttttaactgtggcaacagcgcgcgtgtgtcgtccgcgccgcgcgtcgagccttccc	420
D	b	546	CTCTTCCAGTGGGCCGTGAGCCGTGCATGTGCTCGCCACCCGCCCTGCATGTCC	605
Q	y	421	caaggcacctcagaacgagacgctatgcaagacatgtgcacygcgacacattcaagc	480
D	b	606	CGAGGCGCCATCAGACAGGAGCCGCTGTGAAGCCACCTGTGCGCGCCGCGAGTCCCTTCAACC	665
Q	y	481	ctggcccgaggagcctgpcgcctccatacagccacactggctgtcaaggaggctgctcttc	540
D	b	666	CTGGCCCGGAGGCGCTGCCGCTCTTACAGCCACGCTGTGCTCAAGAGAGTGGCGTTCTTC	725
Q	y	541	aactctcagtgctctaccctgctcaqgcaccccgctccaacctgcgcatctgctcac	600
D	b	726	AACCTGCAAGTGTCTTACCCGCTGCTCAGCAGCCGCGCTCAACCTACCATCTGTGAC	785
Q	y	601	ctggtgcygacccccggggccgtgtgcgctccgggagggcgggcccgatactggca	660
D	b	786	CTGGTGCGCAGCCCGCGGCCCTGCTGGCGCTCCGGGAGCAGACAGCCAAAGGCTCTGGCG	845
Q	y	661	cgcgacaacgcatcgtctagggaaccaacgcaagtgtgtgagtgagcgacacctcaactg	720
D	b	846	CGTGACAACGGCATCTGCTGTGGGCACCAACGGCACGTGGGTGGAGCGCGACCCCGGCGTG	905
Q	y	721	cgcctgattcgcgaggtgtgcgcgagccacgtgcgcatgcgcgaggcgcccaactcaag	780
D	b	906	CGCGTGTGCGCGAGGTGCGGTAGCCACCGTAGCGATCGCGGAGCGCGCACACTCAAG	965
Q	y	781	cscgcaacctcctgcgcgcgcctaccgcctgtgcctctcagagacctggcgcgagg	840
D	b	966	CAGCCACCCCTTCTCGCGGGCCGCTTACCCTCCCTGGTGTGCTTCGAGGACCTTGGCGGGAG	1025
Q	y	841	cgcgtgcgagagatccgcgcactctacgcttccacggcctgacctcagccacacagctc	900
D	b	1026	CGCGTGGCAGAATCCGTCCGCTCTTACGCGCTTACGTGGGTTCATCTCAGGCCACAGCTC	1085
Q	y	901	gaggcctggatccacaacatcacccacgggtgcgggatcsgcaagccaaatcgaggccttc	960
D	b	1086	GAGGCTTGATCCATAACATCACCCACGGATCTGGACCTGGTGGCGCCCGCAGAGCGCTTC	1145
Q	y	961	cbaactctctaggaaatgcgcgaacdtctcccaggccttggcccaacgcyttgccttc	1020
D	b	1146	AAGACTCTGTCAGGAATCGCTCCGCTCTTACGCGCTTACGTGGGTTCATCTCAGGCCACAGCTC	1205
Q	y	1021	actaagatcctgcgctgcagaggtgtggcccgcgcgctgcagctgctgggctaccgg	1080
D	b	1206	GCCAAGATCCGCCGCTCAGGAACCTGCGCTGTCGCTGTCAGCTGCTGCGGTACCGGTACCGS	1265
Q	y	1081	cctgtgtaactctgggacacagcagcgtgacctcaacctggatatgtgtgcacagagcc	1140
D	b	1266	CCTGTGTACTCTGAGAGCAGCAGCGCAACTCSCCCTTGATCTGTGTGCTGCCACGAGGC	1325
Q	y	1141	ccagacacacttcagctgggcatc	1163
D	b	1326	CTGAACGGCTTCACTTGGGCATC	1348

```

RESULT 10
AC026419/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AC026419 159072 bp DNA HTG 20-APR-2001
Homo sapiens chromosome 5 clone CTD-2113H21, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
AC026419
AC026419.3 GI:13699674
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 159072)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 159072)
DOE Joint Genome Institute.
Direct Submission
Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2680 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7711893.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 671130, H460
Center clone name: CITB-HL_2113H21
-----
Summary Statistics
Consensus quality: 140723 bases at least Q40
Consensus quality: 148960 bases at least Q30
Consensus quality: 151290 bases at least Q20
Estimated insert size: 98000; pulse field gel estimation
Estimated insert size: 157572; sum-of-ctontigs estimation
Quality coverage: 7.39 in Q20 bases; pulse field gel estimation
Quality coverage: 4.59 in Q20 bases; sum-of-ctontigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1084: contig of 1084 bp in length
* 1085 1184: gap of unknown length
* 1185 2395: contig of 1211 bp in length
* 2396 2495: gap of unknown length
* 2496 4280: contig of 1785 bp in length
* 4281 4380: gap of unknown length
* 4381 6158: contig of 1778 bp in length
* 6159 6258: gap of unknown length
* 6259 8612: contig of 2354 bp in length
* 8613 8713: gap of unknown length
* 8713 11625: contig of 2913 bp in length
* 11626 11726: gap of unknown length
* 11726 13256: contig of 1531 bp in length
* 13257 13357: gap of unknown length
* 13357 16479: contig of 3123 bp in length
* 16480 16579: gap of unknown length
* 16580 19343: contig of 2764 bp in length
* 19344 23596: contig of 4153 bp in length
* 23597 23696: gap of unknown length
* 23697 29181: contig of 5485 bp in length
* 29182 32882: gap of unknown length
* 32882 35079: contig of 5798 bp in length
* 35080 35179: gap of unknown length

```


QY 481 ctggccgggggagcgtgcgtctctacagcaaacgtgtgtctcaaggaggtgcgtctctt 540
 Db 584 ctggctcaggagcgtgcagctctctatagtcacgtctgtctcaggaggtgcgtctctt 643
 QY 541 aacctgcaggtgctctaccgctctcagcagcccgctcaacctgcagcagcgtgcac 600
 Db 644 AACCTACAGGTGCTCTACCGCTGCTCAGGACCGCTCAACCTGCGCATCGTGGAC 703
 QY 601 ctggtgcagcagcggcgccgctgctgcgtctcccgaggcgccggcgccgatactggca 660
 Db 704 CTAGTGCAGCAGCGCGCGCGCTGCTCGCTCCCGAGAGCAGACGCAAGGCGCTGGCA 763
 QY 661 cgcgcaacgcgcagctgctgctggcaccagcgcaagtgggtgagggcgagccctcaccg 720
 Db 764 CGGACAAATGGCATGCTCTCTGGTACCAACGGCAGCTGGGTGGAGCGCGACCCCGGCTG 823
 QY 721 cgcctgattcgcagagtgctgcgcagcagcagctgcgcagcagcagcagcagcag 780
 Db 824 CGCGTGGTCAACGAGGTATGCCGAGCCATGTGGCAATCGCAGGAGCGCTTGGCAAG 883
 QY 781 cgcgcacctctctgcgcgcgctaccgctgctgctgctgctgctgctgctgctgctg 840
 Db 884 CGCGCGCCTCTCTTCAAGATCGTACGGCTGCTGGTGGTGGTGGTGGTGGTGGTGG 943
 QY 841 cgcgtggcagagatcgcgcagcagcagcagcagcagcagcagcagcagcagcagc 900
 Db 944 CCACATCACCGTAATCCGTGAATCTATGCTCTACCGCGCTGGTGGTGGTGGTGGTGG 1003
 QY 901 gaggcctggatccacacatcaccacacgggtcggggctgcgcagcagcagcagcagc 960
 Db 1004 CAGACTTGGATCCACAATATCAGCATGGTTCAGGCGCCAGCGCGCGCGCTTGCACAG 1063
 QY 961 catacttgctaggaatgcgcagcagcagcagcagcagcagcagcagcagcagcagc 1020
 Db 1064 AAGACCATCCAGGAGTGGCTCAGTCTATCCAGGCTGGCGCCAGCAGCTGGCTGGCT 1123
 QY 1021 actaagatcctgcgcgtcagaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
 Db 1124 GCCAAGATTCGCCGGTCCAGAACTGTGGGGGGTGCACCTGCTGGTGGTGGTGGTGG 1183
 QY 1081 cctgtgactctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1140
 Db 1184 TCTGTGCTATTGGAGCTTTCAGCAAGGAGGACTCTCTGTGACCTCTCTGTGACCTCTCT 1243
 QY 1141 ccagaccacttcagctgggagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1166
 Db 1244 ANGGACAGTTCAAGTGGGCTATGTC 1269

RESULT 14
 AF131235 2032 bp mRNA PRI 10-AUG-1999
 LOCUS Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,
 complete cds.
 ACCESSION AF131235
 VERSION AF131235.1 GI:4927113
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2032)
 Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R.,
 Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S.
 Sulfotransferases of two specificities function in the
 reconstitution of high endothelial cell ligands for L-selectin
 J. Cell Biol. 145 (4), 899-910 (1999)
 99264336
 10330415
 2 (bases 1 to 2032)
 Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y.,
 Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and
 Hemmerich,S.

TITLE

JOURNAL

Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA

FEATURES

source

Location/Qualifiers

1..2032

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="high endothelial"

/tissue_type="tonsil"

182..1342

/function="biosynthesis of endothelial ligands for
 L-selectin"

/note="HPC-GlcNAc-6-S"

/codon_start=1

/product="N-acetylglucosamine 6-O-sulfotransferase"

/protein_id="AAB3015.1"

/db_xref="GI:4927114"

/translation="MLPKMKLLFLVSOMAILLFPHMYSHNLSMKAOERMH
 VLSSRSRSGSFVGLFQHDVFLMEPAWHVMTFKQSTAMLMHVRDLTRAVF
 LCDMSVDFAYMEPRGPRSSLPFOWNSRALCSAPACDIPODEIPRAHRLCSOP
 FEVEKACRSYSHVLPKVRPNLOSPLKDPNLNHLVHLVDRPDRSRRTK
 GDLMDSRVMSGHEOKLKKEDOPXYVMOVICQSOLELYKTLQSLPKALQERVLLRY
 EDLARPAVATSRMYTFEVLPLHLQITWVINITGKMGDMHETHWARDNINVSQAW
 RWSLPYKVSRLQKACGDAMNLLGTRHVRSEQERNLLLDLLSTWTFPEQLH"

CDS

BASE COUNT 468 a 569 c 490 g 505 t

ORIGIN

Query Match

34.1%; Score 400; DB 9; Length 2032;

Best Local Similarity 63.6%; Pred. No. 3.3e-49;

Matches 645; Conservative 0; Mismatches 360; Indels 9; Gaps 2;

QY 118 gatcgtgacagtgcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 177

Db 302 GAGCGCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361

QY 178 ctcttcagcagcagcccgagctctctacgtgatgagcccgctgcatgtatgagcc 237

Db 362 CTCTTTGGGAGGACCCAGAGTGTCTACCTGATGGAGCCCGCTGGCAGCTGTGGATG 421

QY 238 accctgtcagggcagcgcggaacgctgcacatggcgtgctgcgcagcctgatgcgtct 297

Db 422 ACCTTCAAGCAGACGACCGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481

QY 298 atctttgtgcacatggagcgtgttgatgctcctacatccacagag ---ccgaaacctg 354

Db 482 GTCTTCTGTGGCAGACGACGCTGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541

QY 355 tcgcctttttcaactgggcaacgagcgcgcgtgtgctgcgcgcgcgcgcgcgcgcgcgcgc 414

Db 542 TCCAGCCTCTTTCAGTGGGAGAAACAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 601

QY 415 ttcccgagcagcaccatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474

Db 602 ATCCACAAAGATGAATCATCCCGCGGCTCATCTGAGGCTCCTGTGTCAGTCAACAGCCC 661

QY 475 ttcagcctggcccgaggcgtgcgcctcctcagcagcagcagcagcagcagcagcagcagcag 534

Db 662 TTTGAGGTGTGGAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721

QY 535 tcttcaacctcaggtgtctaccgcgtcgtcagcagcagcagcagcagcagcagcagcagcag 594

Db 722 TTCTTCAACCTGAGTCCCTCTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781

QY 595 gtgcacctgtgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 654

Db 782 GTGACCTGTGCGGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841

QY 655 ctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 814

Db 842 CTCATGATTGACATCGCATGTGTGATGGGCGACCATGACGACAAACTCAAGAGCAGGAC 901

QY 715 cactgcgcctgattcgcagaggtgtgtccgcagcagcagcagcagcagcagcagcagcagcag 774

Fri Feb '1 20:20:55 2002

Search completed: January 31, 2002, 18:41:36
Job time: 12745 sec

us-09-593-828-4.rge

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:45:39 ; Search time 21.68 Seconds
(without alignments)
404.810 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRSSKTVTVLLAQTF.....LTDLVLRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 21252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1008	49.1	386	4	US-09-045-284A-2
2	665	32.4	483	3	US-09-263-023-2
3	656.5	32.0	484	3	US-09-263-023-4
4	577.5	28.2	479	2	US-08-899-514-2
5	540.5	26.4	458	2	US-08-655-878-2
6	105	5.1	359	3	US-09-150-133-11
7	105	5.1	359	3	US-09-150-141-11
8	105	5.1	359	4	US-09-374-493-11
9	105	5.1	359	4	US-09-374-824-11
10	105	5.1	359	4	US-09-374-492-11
11	90.5	4.4	380	3	US-09-150-133-9
12	90.5	4.4	380	3	US-09-150-141-9
13	90.5	4.4	380	4	US-09-374-493-9
14	90.5	4.4	380	4	US-09-374-824-9
15	90.5	4.4	380	4	US-09-374-492-9
16	90.5	4.4	384	2	US-08-673-789-9
17	86.5	4.2	376	3	US-09-150-133-7
18	86.5	4.2	376	3	US-09-150-141-7
19	86.5	4.2	376	4	US-09-374-493-7
20	86.5	4.2	376	4	US-09-374-824-7
21	86.5	4.2	376	4	US-09-374-492-7
22	86.5	4.2	377	3	US-09-150-133-5
23	86.5	4.2	377	3	US-09-150-141-5
24	86.5	4.2	377	4	US-09-374-493-5
25	86.5	4.2	377	4	US-09-374-824-5
26	86.5	4.2	377	4	US-09-374-492-5
27	86.5	4.2	566	2	US-08-484-993B-41

ALIGNMENTS

RESULT 1

US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192

; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette

; APPLICANT: Rosen, Steven D.

; APPLICANT: Hemmerlich, Stefan

; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

; FILE REFERENCE: 6510-107US1

; CURRENT APPLICATION NUMBER: US/09/045,284A

; CURRENT FILING DATE: 1998-03-20

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 386

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-045-284A-2

Query Match 49.1%; Score 1008; DB 4; Length 386;

Best Local Similarity 52.3%; Pred. No. 1.1e-102;

Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

28	86.5	4.2	566	2	US-08-484-158B-41	Sequence 41, Appl
29	86.5	4.2	566	2	US-08-484-596A-41	Sequence 41, Appl
30	86.5	4.2	566	2	US-08-480-150A-41	Sequence 41, Appl
31	86.5	4.2	566	3	US-08-458-731-41	Sequence 41, Appl
32	86.5	4.2	566	3	US-08-149-223A-41	Sequence 41, Appl
33	86.5	4.2	1711	3	US-08-369-822C-10	Sequence 10, Appl
34	86.5	4.2	1711	3	US-08-582-776C-10	Sequence 10, Appl
35	86.5	4.2	1711	3	US-08-434-831B-10	Sequence 10, Appl
36	83.5	4.1	390	4	US-09-108-020-36	Sequence 36, Appl
37	83	4.0	15281	2	US-08-471-119A-2	Sequence 2, Appl
38	81.5	4.0	829	4	US-08-444-818-69	Sequence 69, Appl
39	81.5	4.0	1786	4	US-08-444-818-54	Sequence 54, Appl
40	81.5	4.0	2261	4	US-08-444-818-66	Sequence 66, Appl
41	81.5	4.0	2436	4	US-08-444-818-75	Sequence 75, Appl
42	81.5	4.0	2772	4	US-08-444-818-89	Sequence 89, Appl
43	81.5	4.0	2894	4	US-08-466-975A-23	Sequence 23, Appl
44	81.5	4.0	2894	2	US-08-391-671A-23	Sequence 23, Appl
45	81.5	4.0	2894	3	US-08-467-902A-23	Sequence 23, Appl

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/899,514
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DANIEL E ALTMAN
 ; REGISTRATION NUMBER: 34,115
 ; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 714 760 0404
 ; TELEFAX: 714 760 9502
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 479
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-899-514-2

Query Match 28.2%; Score 577.5; DB 2; Length 479;
 Best Local Similarity 35.7%; Pred. No. 4.6e-55;
 Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps 8;

QY 32 PSSPAGGEDRVHVLSSWRSGSLGQLFSQHPDVVYLMPEAHVWTTLS----QGSAA 87
 DB 123 PPPAVAGERRHVLMMATRTGSSVGEFFNQGNIFILFEPLMHIERTVTFEGGANVCSAL 182
 QY 88 TLMAYRDLMSIFLDCMDVDFAY--MPOSNLSAFNFWATSRALCSPPACSPFRGTI 144
 DB 183 GSALVYEDVLKFLCDIVLHFFITPDPDLTOPMFRGSSRLCEDPVCYTFVKKVF 242
 QY 145 SKDVKCTLTCTQPTSLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVVD 204
 DB 243 EKYHCKNRRCGLPNVTIAAECRKHMAKAVRIQLEFLQPLAEDPDLRLVQLVRD 302
 QY 205 PRAVLRSRAAGPILARONGIVLG--TINGKWVE-----ADPHRLRLREVCSSHVRIA 254
 DB 303 PRAVLASRWA-----FAGKYKTWKWLDDGQDGLREVEVQRLGNCS-IRLS 351
 QY 255 EATLPPPLRGYMLVRFEDLAREPLAEIRALYAFGLTLTPOLEAWIHTH----G 310
 DB 352 AELGLQPAWLRGYMLVRYEDVARGPLQKAREMYPFAGIPLTPQVEDWIOKNTQAAHDG 411
 QY 311 SGICKPIEAPHSSRNARNVQAWRHLPFTKILRVQEVCAQALQLLGYRVPYSADQORD 370
 DB 412 SGI-----YSTQKNSSEQFEKWRFSMPFKLAQVQAPCGPAMFLFGYKLARDAALTN 464
 QY 371 LTLDLVLPR 380
 DB 465 RSVSLLEERG 474

RESULT 5
 US-08-655-878-2
 ; Sequence 2, Application US/08655878
 ; Patent No. 5827713
 ; GENERAL INFORMATION:
 ; APPLICANT: FUKUTA, MASAKAZU
 ; APPLICANT: HABUCHI, OSAMI
 ; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE:
 ; STREET:

; CITY:
 ; STATE:
 ; COUNTRY:
 ; ZIP:
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/655,878
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME:
 ; REGISTRATION NUMBER:
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE:
 ; TELEFAX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 458
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-655-878-2

Query Match 26.4%; Score 540.5; DB 2; Length 458;
 Best Local Similarity 34.3%; Pred. No. 5.1e-51;
 Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;

QY 36 AGEDRVHVLSSWRSGSLGQLFSQHPDVVYLMPEAHV--WTTLSGSAATLHMAY 93
 DB 107 AAFEPHRRHVLMMATRTGSSVGEFFNQGNIFILFEPLMHIERTVTFEGGANVCSAL 166
 QY 94 --RDLMSIFLDCMDVDFAY--MPOSNLSAFNFWATSRALCSPPACSPFRGTISKQD 148
 DB 167 VYRDVLQQLLGLDYLIFESFISPAPEEHLTAALFRGSSHSLSCEPVCYTPSLKVVFKYH 226
 QY 149 VCKTLCTROPFSLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVROPRAV 208
 DB 227 CXNRRCGLPNVTIAAECRKHMAKAVRIQLEFLQPLAEDPDLRLVQLVRDPRV 286
 QY 209 LRSREAGPILARONGIVLGTNGKWE-----ADPHRLRLREVCSSHVRIAETATL 259
 DB 287 LVSRMVA-----FSGKYESWKWAAEGEAPLOEDEVOVLRGNCS-IRLSAELGL 335
 QY 260 KPPFLRGYRLVRFEDLAREPLAEIRALYAFGLTLTPOLEAWIHTHITGSGICKPIEA 319
 DB 336 RQPRWLGRYMLVRYEDVARAPLRKALEMYRFAGIHPPTQVEEWIRANTQAP---QDSNG 392
 QY 320 FHTSSRNARNVQAWRHLPFTKILRVQEVCAQALQLLGYRVPYSADQORDLTLDLVLPR 379
 DB 393 IYSTQKNSSEQFEKWRFSIPFKLAQVQDACEPAMRLFGYKLASSAGELNRSLSL-LEE 451
 QY 380 GP 381
 DB 452 GP 453

RESULT 6
 US-09-150-133-11
 ; Sequence 11, Application US/09150133B
 ; Patent No. 6060295
 ; GENERAL INFORMATION:
 ; APPLICANT: The Board of Regents of the University of Oklahoma
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF

```

; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150.133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-11

Query Match
Best Local Similarity 5.1%; Score 105; DB 3; Length 359;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VVILLAAQTTCILLFI-----ISRPGSSPAGGEDRVH-----VLY 46
Db 31 IYIFFCFTICLLIFSSIKCKKIQEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGOLFSDPDV-----FYLMPEAW-HVWTTLSQGSAAATLHMAYR 94
Db 91 GGVPESGTTLMRAILDADHPDVRGCGETMLLPSPFLTQAGWRNDWV---NNSGIT- 141
QY 95 DLMRSIFLCMDVDFDAYMPQSRNLSAFFNWATSRALCSPACSAFPGTISK-ODVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTROPFSLAREACKSYSHVVLKEVRFFNLOVLYPLSDPALNLRIVHLVDRPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPTR-----RLYP-----NAKFLMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILREVCRSHVRIAEAAATLKPPPLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NQELRKMTFOCNN---APGQCIK-----V 247
QY 273 RFEDLAREPLAEIRALYAFGTGLTLPQLEAWIHNTHGSGIGKPI-----EAEHTSS-RNA 327
Db 248 YIERLIQKPAEILRTITNFDLPFSQOM-----LRHODLIGDEVLDNDQEFSSASOVKNS 301
QY 328 RNVS--QAWRHALPFTKILRVOEVCAGALOLLGY-----RPVYS 364
Db 302 INTKALTSWDFCSEETLRKLDV-APFLGILGYDTSISKPDYS 344

```

RESULT 8

```

US-09-374-493-11
; Sequence 11, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374.493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150.133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-493-11

```

Query Match 5.1%; Score 105; DB 4; Length 359;

Best Local Similarity 19.8%; Pred. No. 0.0038;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

```

QY 11 VVILLAAQTTCILLFI-----ISRPGSSPAGGEDRVH-----VLY 46
Db 31 IYIFFCFTICLLIFSSIKCKKIQEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGOLFSDPDV-----FYLMPEAW-HVWTTLSQGSAAATLHMAYR 94
Db 91 GGVPESGTTLMRAILDADHPDVRGCGETMLLPSPFLTQAGWRNDWV---NNSGIT- 141
QY 95 DLMRSIFLCMDVDFDAYMPQSRNLSAFFNWATSRALCSPACSAFPGTISK-ODVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTROPFSLAREACKSYSHVVLKEVRFFNLOVLYPLSDPALNLRIVHLVDRPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPTR-----RLYP-----NAKFLMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILREVCRSHVRIAEAAATLKPPPLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NQELRKMTFOCNN---APGQCIK-----V 247

```

RESULT 7

```

US-09-150-141-11
; Sequence 11, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150.141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-141-11

```

Query Match 5.1%; Score 105; DB 3; Length 359;

Best Local Similarity 19.8%; Pred. No. 0.0038;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

```

QY 11 VVILLAAQTTCILLFI-----ISRPGSSPAGGEDRVH-----VLY 46
Db 31 IYIFFCFTICLLIFSSIKCKKIQEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90

```

```
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; ORGANISM: Caenorhabditis elegans
US-09-374-824-11

Query Match          5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0038;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VVILLAAQTTCILLFI-----ISRPGSPSAGGEDRVH-----VLV 46
Db 31 IYIFCFCTICLLIFSSICKKLOEKLSKESLIFNEQDARHSRRLSNLEQLIFV 90
QY 47 LSSWRSGLQFLQFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTILMRAILDAHPDVRCGGETMLPSFLTWQAGWRNDWV---NNSGIT--- 141
QY 95 DLMSIFLCMDVDYDAYMPOQRNLSAFNNWATSRALCPSPACSAFPGRTISK-ODVCKTL 153
Db 142 -----QEVFD-----DAVSAFITIVAKHSELAPRL 167
QY 154 CTROPFLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDPRAVLSR- 212
Db 168 CNKDP-----YTALWLPITR-----RLYP-----NAKFLMIRCARAVVHSMI 205
QY 213 EAAGPILARDNGIVLTGCKWVEADPHLRIREVCRSHVRVIAEAATLKPPFLRCRYLV 272
Db 206 ERKVFNAGYNTSDEISMFVQW---NOELRKMFTQCNN---APGCCIA-----V 247
QY 273 RFEDLAREPLAERIALYAFGTLTLPQLEAMHNIHSGIGKPI-----EAFHTSS-RNA 327
Db 248 YVERLIQKPAEELIRITNFDLPFSQOM-----LRHODLIGDEVLDNQEFASQVKN 301
QY 328 RNVS--QAWRHALPFTKILRVQEVCAQALQLLY-----RPVYS 364
Db 302 INTKALTSWDFCFSEETLRKLDV-APFLGILGYDTSTISKPDYS 344

RESULT 11
US-09-150-133-9
; Sequence 9, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-9
```

```
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; ORGANISM: Caenorhabditis elegans
US-09-374-492-11

Query Match          5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0038;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VVILLAAQTTCILLFI-----ISRPGSPSAGGEDRVH-----VLV 46
Db 31 IYIFCFCTICLLIFSSICKKLOEKLSKESLIFNEQDARHSRRLSNLEQLIFV 90
QY 47 LSSWRSGLQFLQFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMAVR 94
Db 91 GGVPRSGTILMRAILDAHPDVRCGGETMLPSFLTWQAGWRNDWV---NNSGIT--- 141
QY 95 DLMSIFLCMDVDYDAYMPOQRNLSAFNNWATSRALCPSPACSAFPGRTISK-ODVCKTL 153
Db 142 -----QEVFD-----DAVSAFITIVAKHSELAPRL 167
QY 154 CTROPFLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDPRAVLSR- 212
Db 168 CNKDP-----YTALWLPITR-----RLYP-----NAKFLMIRCARAVVHSMI 205
QY 213 EAAGPILARDNGIVLTGCKWVEADPHLRIREVCRSHVRVIAEAATLKPPFLRCRYLV 272
Db 206 ERKVFNAGYNTSDEISMFVQW---NOELRKMFTQCNN---APGCCIA-----V 247
QY 273 RFEDLAREPLAERIALYAFGTLTLPQLEAMHNIHSGIGKPI-----EAFHTSS-RNA 327
Db 248 YVERLIQKPAEELIRITNFDLPFSQOM-----LRHODLIGDEVLDNQEFASQVKN 301
QY 328 RNVS--QAWRHALPFTKILRVQEVCAQALQLLY-----RPVYS 364
Db 302 INTKALTSWDFCFSEETLRKLDV-APFLGILGYDTSTISKPDYS 344

RESULT 11
US-09-150-133-9
; Sequence 9, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-9
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:20:51 ; Search time 4771.45 Seconds
(without alignments)
7067.083 Million cell updates/sec

Title: US-09-593-828-3

Perfect score: 2044

Sequence: 1 ggctcaggctccactgtgtc.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo_hum.*
- 31: em.htgo_inv.*
- 32: em.htgo_rod.*
- 33: em.htg_hum.*
- 34: em.htg_inv.*
- 35: em.htg_rod.*
- 36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1973	96.5	2170	9	AF176838	AF176838 Homo sapi
2	1774.6	86.8	3278	9	AF246718	AF246718 Homo sapi
3	1753.6	85.8	181566	2	AC025287	AC025287 Homo sapi
4	1742.6	85.3	159072	2	AC026419	AC026419 Homo sapi
5	1719	84.1	71503	9	AF219991	AF219991 Homo sapi
6	1628.8	79.7	179065	2	AC009163	AC009163 Homo sapi
7	1248.6	61.1	1462	9	AF176839	AF176839 Homo sapi
8	1101	53.9	3786	9	AF280086	AF280086 Homo sapi
9	1052	51.5	2544	9	AF219990	AF219990 Homo sapi
10	1046.4	51.2	159072	2	AC026419	AC026419 Homo sapi
11	1046.4	51.2	208185	2	AC009105	AC009105 Homo sapi
12	751.2	36.8	1989	10	AF176840	AF176840 Mus muscu
13	750.4	36.7	1740	10	AF176841	AF176841 Mus muscu
14	568	27.8	215647	2	AC068591	AC068591 Homo sapi
15	556.8	27.2	174187	2	AP002792	AP002792 Homo sapi
16	542.6	26.5	137499	2	AC015931	AC015931 Homo sapi
17	524.8	25.7	165762	2	AP001582	AP001582 Homo sapi
18	461.2	22.6	179065	2	AC009163	AC009163 Homo sapi
19	400	19.6	2032	9	AF131235	AF131235 Homo sapi
20	398.4	19.5	1333	9	AF149783	AF149783 Homo sapi
21	398.4	19.5	1992	9	AF280088	AF280088 Homo sapi
22	398.4	19.5	2011	9	AK026635	AK026635 Homo sapi
23	398.4	19.5	183133	2	AC010547	AC010547 Homo sapi
24	358.8	17.6	2201	10	AF109155	AF109155 Mus muscu
25	354	17.3	1926	10	AF131236	AF131236 Mus muscu
26	235.2	11.5	2156	6	AR071396	AR071396 Sequence
27	235.2	11.5	2156	6	EL4937	EL4937 Human mRNA
28	235.2	11.5	2731	9	AB012192	AB012192 Homo sapi
29	235.2	11.5	6961	9	AB017915	AB017915 Homo sapi
30	235.2	11.5	182677	2	AC073370	AC073370 Homo sapi
31	235.2	11.5	196465	9	AC022392	AC022392 Homo sapi
32	231	11.3	176752	2	AL590620	AL590620 Homo sapi
33	228.8	11.2	172176	2	AL450304	AL450304 Homo sapi
34	226.2	11.1	103567	9	HS124C6	HS124C6 Homo sapi
35	224.4	11.0	54666	9	AC073487	AC073487 Homo sapi
36	223	10.9	154952	2	AL358153	AL358153 Homo sapi
37	223	10.9	166820	2	AL590126	AL590126 Homo sapi
38	222	10.9	140210	2	AC002993	AC002993 Homo sapi
39	221.6	10.8	195382	2	AL445254	AL445254 Homo sapi
40	220.8	10.8	151068	9	AL133405	AL133405 Homo sapi
41	220.6	10.8	148463	2	AC012140	AC012140 Homo sapi
42	220.6	10.8	207408	2	AC092872	AC092872 Pan trogl
43	220.6	10.8	213227	2	AC021027	AC021027 Homo sapi
44	220.4	10.8	57573	2	AC087509	AC087509 Homo sapi
45	220.4	10.8	182012	2	AC025590	AC025590 Homo sapi

ALIGNMENTS

RESULT 1

AF176838

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2170)

Lee, J.K., Bhakta, S., Rosen, S.D. and Hemmerich, S.

Cloning and characterization of a mammalian

N-acetylglucosamine-6-sulfotransferase that is highly restricted to

intestinal tissue

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

99423499

2 (bases 1 to 2170)

AF176838 2170 bp mRNA PRI 22-SEP-1999
Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase
(I-GlcNAc-6-ST) mRNA, complete cds.

AF176838.1 GI:5917705

AUTHORS Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
 TITLE Direct Submission
 JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA
 FEATURES Location/Qualifiers
 source
 1. .2170
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q23.1"
 /tissue_type="intestine"
 1. .2170
 /gene="I-GlcNAc-6-ST"
 344. 1516
 /gene="I-GlcNAc-6-ST"
 /note="possibly involved in biosynthesis of mucosal
 ligands for L-selectin"
 /codon_start=1
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 /protein_id="AAD56000.1"
 /db_xref="GI:5917706"
 /translation="MWLPFRSSKTVIVLLIAQTLLILFIISRPSPAGDERVHV
 IVLSSWAGSFLQLESQHPEDVYLMPEAWHTLSQGSAAITLHMAVRDLMSIFL
 CDMVFDAYMPSQSNLSAFNWAISRALCSPPACSAPRWTISKDVKLTCTROPFS
 LARACRSYSHVLEKEVFFNLQILPILSDPALNLRIVHLVROPRAYLRSREAGPI
 LARDNGIVLGNKWEADPHRLIREVCRSHVRAEALIKPPFLGRYELVFED
 LARPLAIRALYAFGLTLTPQLFAMWHTHSGIGKPIEAFHTSSRNARNWSQAW
 RHAPLEKILRVQEVCAALDILGYRPVYSADQQRDLTLDLVLRPDPHFWSAPD"
 449 a 682 c 612 g 427 t
 BASE COUNT
 ORIGIN
 Query Match 96.5%; Score 1973; DB 9; Length 2170;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1998; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
 QY 41 gaaagctccagtgaggaaactgtctcttgagactctgtgtggcatagatgaattcaac 100
 Db 167 GAAAGCTCCAGTGGAGAACTGCTCTCTGGAGACTCTGTGTGGCATAGAGTGAATCAAC 226
 QY 101 caacttaagaagacctgtgtcttctggaacacagatgtcgagacatctcccatgatt 160
 Db 227 CACCTTAAGAAGACCTCTGTGCTTCTTCCTGGAAACACAGATGTCGAGACATCTCCATGATT 286
 QY 161 tttgaatcagctgtgacgtctccagcagccctggaagctgtgccccagccgcgcgatg 220
 Db 287 TGTGATCAGGTTGACGCTCTCCAGAGCCCTGACGCTGGCCGCCAGCCGCCGCGCATG 346
 QY 221 tggctgcaaggtttctccagcaagacagtgacagtgctctctctctgtgacacacacatgc 280
 Db 347 TGGCTGCCACGGTCTCCAGCAAGACAGTGAAGTGTCTCTGTCCTGGCAGACACACCTTGC 406
 QY 281 ctctgtcttctcaatctccagccagggccctcatctccctctgtgacacacacacatgc 340
 Db 407 CTCTGTCTCTCAATCTCCCGGCCAGGCCCTCATCCAGCCGCCGCCGAGGATCGT 466
 QY 341 gtgcacgtgtgtgtgtctctgtggtgctggtgctctctctctctctctctctctctct 400
 Db 467 GTGCACGT 526
 QY 401 agccagcaccocagcgttctctacatgtatggagccgagtgatgtgtggaccacccgtg 460
 Db 527 AGCCAGCACCCCGAGCTTCTACCTGATGGAGCCGCGTGGCATGTGTGGACACCCG 586
 QY 461 tgcagggcagcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 520
 Db 587 TCGAGGGCAGCGCGCAGACCGTGCACATGGCGGTGCGGACCTGATGCGCTATCTT 646
 QY 521 ttgtgcacatggacgtgtttgatgctacatgcccacagagccagcagcagcagcagcagc 580
 Db 647 TTGTGCGACATGGACGT 706
 QY 581 ttcaactgggcaacagagccgcgcgtgtgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 640

Db 707 TTCAACTGGGCACAGAGCCGCGCGCTGTGCTCGCGCCGCGCTGCAGCGCTTTCGCCGA 766
 QY 641 ggcaccatcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 700
 Db 767 GGCACCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 826
 QY 701 gccgggagc 760
 Db 827 CCCCAGGAGGCTGCGCGCTCTACAGCCACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 886
 QY 761 ctgcaagtgctctaccgc 820
 Db 887 CTGCAAGTGTCTTACCGCTGCTCAGCGGACCCCGGCTCAACCTGGCATGTGTGTGTGT 946
 QY 821 gtgc 880
 Db 947 GTGCGGACCCCGCGCGCTGTGCGCTCCCGGAGCGCGCGCGCTACTTGGCAGCG 1006
 QY 881 gacaacggcagctgt 940
 Db 1007 GACAACGGCATGCTGTGTGGGCACCAACGCAAGTGGTGTGTGTGTGTGTGTGTGTGTGT 1066
 QY 941 ctgattcgcaggt 1000
 Db 1067 CTGATTGCGGAGGT 1126
 QY 1001 ccaccttctgtgc 1060
 Db 1127 CCACCTTCTGTGCGCGCGCTACCGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1186
 QY 1061 ctggcagagatccgc 1120
 Db 1187 CTGSCAGAGATCGCGCACTCTAGCGCTTTCACCGGCTGTGTGTGTGTGTGTGTGTGTGT 1246
 QY 1121 gctgtgtac 1180
 Db 1247 GCCTGTGATTCACACATCACCCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1306
 QY 1181 actctgtctagaaatgc 1240
 Db 1307 ACTTCTGT 1366
 QY 1241 aagatcctgc 1300
 Db 1367 AAGATCTCTGCGCGT 1426
 QY 1301 gtgtactctgc 1360
 Db 1427 GTGTACTCTGTGCGACACAGCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1486
 QY 1361 gaccacttcagctgt 1420
 Db 1487 GACCATTCTGT 1546
 QY 1421 tgtgtgtgc 1480
 Db 1547 TGT 1606
 QY 1481 a-caggtcctactatcaacccggaggtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1539
 Db 1607 AGCAGTCTCTGT 1666
 QY 1540 acgtttcttct 1599
 Db 1667 ACGTCTTCT 1726
 QY 1600 caggtgcactcttcaagaaac 1659
 Db 1727 CAGGTGCACCT 1785
 QY 1660 agttactgtcaaatataatataatgtgtgcagccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1719

Db	1786	AGTTACTGCTAAATTAATTAATTAATGTTGTCAGGCGGGTGGTGGTCTATGCTGTAA	1845
QY	1720	tccagcatttgagagcctgagcggttgatcacctcaggtcaggattcaaacacagc	1779
Db	1846	TCCAGCATTGAGAGCTGAGCGGGTGGATCACTGAGTCAGGATCAAAACCCAGC	1905
QY	1780	ctggccaacatagtgaaacccctctctactcaaaatgcaaaaattagtcggcgctggtg	1839
Db	1906	CTGCGCAACATAGTGAACCCCTCTCTACTAAAATGCAAAAATTAAGTCGGCGTGGTG	1965
QY	1840	gcacactcctgtaaccagactactaggagctgagtggtgggaaatcaccttggaactcca	1899
Db	1966	GCACATCTCTGTAATCCAGCTACTAGGAGCTGAGTGGGAGAAATCACTTGGACTGCCA	2025
QY	1900	aagtgaggtgcagtaagctaaatcatgcacatgcacccctagcttggttggaagc	1959
Db	2026	AAGGTGAGGTTGCAGTAACTGAATCATGCACTGCACCTGACCTTGGTGGCGAAAGC	2085
QY	1960	aaaactctatcaaaaaataattaataattgtccaaagtcctgcgaaaaaataaaaa	2019
Db	2086	AAAACCTATCAAAAAATAATTAATAATTTCTTCAAAAGTCTGCCGAAAAAATAAAA	2145
QY	2020	aaaaaataaaaaaataaaaaaataaaaaa	2044
Db	2146	AAAAAATAAAAAAATAAAAAAATAAAAAA	2170
RESULT	2		
AF246718			
LOCUS	AF246718	3278 bp	31-OCT-2000
DEFINITION	Homo sapiens intestinal GlcNAc-6-sulfotransferase (CHST5) mRNA,	PRI	
ACCESSION	AF246718	complete cds, alternatively spliced.	
VERSION	AF246718.1	GI:11055254	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K.,		
	Nakamura,T., Dots,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,		
	Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A.		
	and Fukuda,M.N.		
TITLE	Macular Corneal dystrophy type I and type II are caused by distinct		
JOURNAL	mutations in a new sulphotransferase gene		
MEDLINE	Nat. Genet. 26 (2), 237-241 (2000)		
PUBMED	20472330		
REFERENCE	11017086		
AUTHORS	2 (bases 1 to 3278)		
	Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K.,		
	Nakamura,T., Dots,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,		
	Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A.		
	and Fukuda,M.N.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAR-2000) Glycobiology Program, The Burnham		
FEATURES	Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
source	Location/Qualifiers		
	1..3278		
gene	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q22: between D16S512 and D16S515"		
	1..3278		
5' UTR	/gene="CHST5"		
	1..1396		
	/gene="CHST5"		
	/note="alternatively spliced"		
	1397..2632		
CDS	/gene="CHST5"		
	/note="I-GlcNAc6ST; alternatively spliced"		
	/codon_start=1		
	/product="intestinal GlcNAc-6-sulfotransferase"		

Fri Feb 1 20:20:49 2002

DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Homo sapiens chromosome 16 clone RP11-490B18, WORKING DRAFT
SEQUENCE, 35 unordered pieces.
AC025287 GI:13786383
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 181566)
DOE Joint Genome Institute.
Direct Submission
Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94596, USA
On Apr 25, 2001 this sequence version replaced gi:11178071.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 594567
Center clone name: RPCI-11_490B18

Summary Statistics
Consensus quality: 153772 bases at least Q40
Consensus quality: 165049 bases at least Q30
Consensus quality: 169147 bases at least Q20
Estimated insert size: 182760; agarose-fp estimation
Quality coverage: 6.17 in Q20 bases; agarose-fp estimation
Quality coverage: 6.33 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1037 ttccagagacctggcgagcgagccgctggcagagatccgcgactctacgccttcacccggc 1096
1096 ttccagagacctggcgagcgagccgctggcagagatccgcgactctacgccttcacccggc 1096
2279 TTCAGAGACTGGCGGGAGCGCTGGCAGAGATCCGGCACTACGCTTCACCGGC 2338
2338 TTCAGAGACTGGCGGGAGCGCTGGCAGAGATCCGGCACTACGCTTCACCGGC 2338
1097 ctgacctcaacgacagctcgagcgctggatccacaaatccacacgagtcgagggatc 1156
1156 ctgacctcaacgacagctcgagcgctggatccacaaatccacacgagtcgagggatc 1156
2339 CTCAGCCTCCGACACAGCTCGAGCGCTGGATCCACAAATCACCACCGGCTCGGGGATC 2398
2398 CTCAGCCTCCGACACAGCTCGAGCGCTGGATCCACAAATCACCACCGGCTCGGGGATC 2398
1157 ggaagcaaatcgagcgcttcactctctaggaatcgcgcaacgcttcacccggc 1216
1216 ggaagcaaatcgagcgcttcactctctaggaatcgcgcaacgcttcacccggc 1216
2399 GGCAAGCAATCGAGCGCTTCCTACTACTCTGCTAGGATCGCGCACTCTCCAGGCC 2458
2458 GGCAAGCAATCGAGCGCTTCCTACTACTCTGCTAGGATCGCGCACTCTCCAGGCC 2458
1217 tggcgccacgcttgcccttcactaaatcctgcgctcgagggagtgcgccggcgcg 1276
1276 tggcgccacgcttgcccttcactaaatcctgcgctcgagggagtgcgccggcgcg 1276
2459 TGGCGCCACGCTTGCCCTTCCTACTAAATCTGCGCTGCGAGAGGTGTCGCGCGCGC 2518
2518 TGGCGCCACGCTTGCCCTTCCTACTAAATCTGCGCTGCGAGAGGTGTCGCGCGCGC 2518
1277 ctgcagctgctggctaccggcctgtactctcgagcagcagcgctgacctcacctg 1336
1336 ctgcagctgctggctaccggcctgtactctcgagcagcagcgctgacctcacctg 1336
2519 CTGCAGCTGCTGGGCTACCGGCTGTGTACTCTCGGACACGACGCTGACCTCACCGTC 2578
2578 CTGCAGCTGCTGGGCTACCGGCTGTGTACTCTCGGACACGACGCTGACCTCACCGTC 2578
1337 gatctggtgctgcacagggccgagaccacttcagctgggcatcgctgactgagaactc 1396
1396 gatctggtgctgcacagggccgagaccacttcagctgggcatcgctgactgagaactc 1396
2579 GATCTGGTGTGCTGCCACGAGGCCACACACCTCAGCTGGGATCGCTGACTGAGAATC 2638
2638 GATCTGGTGTGCTGCCACGAGGCCACACACCTCAGCTGGGATCGCTGACTGAGAATC 2638
1397 tgggcttagagcaagcccgcaactgtgtgctgagccgagcagcagcagcagcagcagc 1456
1456 tgggcttagagcaagcccgcaactgtgtgctgagccgagcagcagcagcagcagcagc 1456
2639 TGGGCTTAGAGCAGCGCCCGCAACTGTGTGCTGCCAGGCCAGGAGCGACTGCTGAGTGG 2698
2698 TGGGCTTAGAGCAGCGCCCGCAACTGTGTGCTGCCAGGCCAGGAGCGACTGCTGAGTGG 2698
1457 aaaagagctggcgagcgatgggaa-caggtccctactatcaacgggagttggggctcc 1515
1515 aaaagagctggcgagcgatgggaa-caggtccctactatcaacgggagttggggctcc 1515
2699 AGAGGAGAGCTGGGGGCGCATGCGGGAACAGGTCCTACTATCAACCGGAGATTGGGGTCC 2758
2758 AGAGGAGAGCTGGGGGCGCATGCGGGAACAGGTCCTACTATCAACCGGAGATTGGGGTCC 2758
1516 tccctgaagtgaagcaagcagcagcttctctctctctctctctctctctctctctctct 1575
1575 tccctgaagtgaagcaagcagcagcttctctctctctctctctctctctctctctctct 1575
2759 TCCCTGAAATGAGGCAAGACTGACGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2818
2818 TCCCTGAAATGAGGCAAGACTGACGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2818
1576 agtctctgagctgct 1635
1635 agtctctgagctgct 1635
2819 AGTCTCTGAGCTGCT 2877
2877 AGTCTCTGAGCTGCT 2877
1636 taccctcttggcgagggagtaagttaagttaagttaagttaagttaagttaagttaagtta 1695
1695 taccctcttggcgagggagtaagttaagttaagttaagttaagttaagttaagttaagtta 1695
2878 TACCTCTTGGCGGAGGAGTAAGTTACTGCTTAATTAATTAATTAATTAATTAATTAAT 2937
2937 TACCTCTTGGCGGAGGAGTAAGTTACTGCTTAATTAATTAATTAATTAATTAATTAAT 2937
1696 ggtgaggtgctgctgctctgtaattccacgactttgagggctgagcggggagatcac 1755
1755 ggtgaggtgctgctgctctgtaattccacgactttgagggctgagcggggagatcac 1755
2938 GGTGCGGTGGCTCATGCTGTATATCCAGCATTTTGAAGAGGCTGAGCGGGTGGATCAC 2997
2997 GGTGCGGTGGCTCATGCTGTATATCCAGCATTTTGAAGAGGCTGAGCGGGTGGATCAC 2997
1756 ctgaggtcagga-ttcaaaaccgagctggcacaatagtaaaacccctctctactaaaa 1814
1814 ctgaggtcagga-ttcaaaaccgagctggcacaatagtaaaacccctctctactaaaa 1814
2998 CTGAGGTCAGGAGTTCGAAACACGCTGGCCACATAGTAAGACCCCTCTCTACTAAAA 3057
3057 CTGAGGTCAGGAGTTCGAAACACGCTGGCCACATAGTAAGACCCCTCTCTACTAAAA 3057
1815 atgcaaaaattagtcggcgctggcgacactctctgtaattccacgactcttaggaggtg 1874
1874 atgcaaaaattagtcggcgctggcgacactctctgtaattccacgactcttaggaggtg 1874
3058 ATGCAAAAATTAGTCCGGCTGGTGGCAGCTCTCTGTATATCCAGCTTATAGGAGGCTG 3117
3117 ATGCAAAAATTAGTCCGGCTGGTGGCAGCTCTCTGTATATCCAGCTTATAGGAGGCTG 3117
1875 aggtgggaaatcacttggactccaaaggtggaggtgaggtgaggtgaggtgaggtgaggtg 1934
1934 aggtgggaaatcacttggactccaaaggtggaggtgaggtgaggtgaggtgaggtgaggtg 1934
3118 AGGTGGGAGAACTACTTGGACTCCAGAGGTGGAGGTTGCATTAAGCTGAGATCATGCCAC 3177
3177 AGGTGGGAGAACTACTTGGACTCCAGAGGTGGAGGTTGCATTAAGCTGAGATCATGCCAC 3177
1935 tgcaccttagcttgggtggcaaaagcaaaactctatcaaaaaataataataataatttgt 1994
1994 tgcaccttagcttgggtggcaaaagcaaaactctatcaaaaaataataataataatttgt 1994
3178 TGCACCTTAGCTTGGGTGGCAGCAAGACTCTATCAAAAATAATAATAATAATTGTT 3237
3237 TGCACCTTAGCTTGGGTGGCAGCAAGACTCTATCAAAAATAATAATAATAATTGTT 3237
1995 caaaagtctgcaaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2035
2035 caaaagtctgcaaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2035
3238 CAAAAGTCCGAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3278
3278 CAAAAGTCCGAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3278

* * *	28152	28251:	gap of unknown length
* * *	28252	30702:	contig of 2451 bp in length
* * *	30703	30802:	gap of unknown length
* * *	30803	34404:	contig of 3602 bp in length
* * *	34405	34504:	gap of unknown length
* * *	34505	37720:	contig of 3216 bp in length
* * *	37721	37820:	gap of unknown length
* * *	37821	40186:	contig of 2366 bp in length
* * *	40187	40287:	gap of unknown length
* * *	40287	42960:	contig of 2674 bp in length
* * *	42961	43060:	gap of unknown length
* * *	43061	46345:	contig of 3285 bp in length
* * *	46346	46445:	gap of unknown length
* * *	46446	50669:	contig of 4224 bp in length
* * *	50670	50769:	gap of unknown length
* * *	50770	53998:	contig of 3229 bp in length
* * *	53999	54098:	gap of unknown length
* * *	54099	58292:	contig of 4194 bp in length
* * *	58293	58392:	gap of unknown length
* * *	58393	64153:	contig of 5761 bp in length
* * *	64154	64253:	gap of unknown length
* * *	64254	71527:	contig of 7274 bp in length
* * *	71528	71627:	gap of unknown length
* * *	71628	78709:	contig of 7082 bp in length
* * *	78710	78809:	gap of unknown length
* * *	78810	87047:	contig of 8238 bp in length
* * *	87048	87147:	gap of unknown length
* * *	87148	94964:	contig of 7817 bp in length
* * *	94965	95064:	gap of unknown length
* * *	95065	102106:	contig of 7042 bp in length
* * *	102107	102206:	gap of unknown length
* * *	102207	108793:	contig of 6587 bp in length
* * *	108794	108893:	gap of unknown length
* * *	108894	117724:	contig of 8831 bp in length
* * *	117725	117824:	gap of unknown length
* * *	117825	132019:	contig of 14195 bp in length
* * *	132020	132119:	gap of unknown length
* * *	132120	153517:	contig of 21398 bp in length
* * *	153518	153617:	gap of unknown length
* * *	153618	181566:	contig of 27949 bp in length.

FEATURES		Location/Qualifiers
Source	1..181566	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="16"
		/clone="RP11-490B18"
BASE COUNT	47714 a 40793 c 39929 g 49716 t	3414 others
ORIGIN		

Query Match	85.8%;	Score 1753.6;	DB 2;	Length 181566;
Best Local Similarity	99.1%;	Pred. No. 5e-281;		
Matches 1795;	Conservative	0;	Mismatches 14;	Indels 3; Gaps

QY	197	cggtggccccccagcgcgcgcagtgtgctgccacgttctccagcaagaactgacagtcg	256
Db	140296	CGAAGCCGCCACGACGGCCGCATGTGGCTGCCATGTTCACAGCAAGACAGTACAGTG	140237
QY	257	ctctctctggcacagaccacctgcctctgtctttctatcatctcccggccaggccctcca	316
Db	140236	CTCCTCTGTGCACAGACCACCTGCTCCTGCTCTTCATCATCTCCCGGCACGGGCCCTCA	140177
QY	317	tccccagccggcgcgaggatcgctgtgcacatgcttgctgtctcttgtgcgtcgcgcgc	376
Db	140176	TCCCCAGCCGGCGGCAGAGATCGTGTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGT	140117
QY	377	taatactcttgggccaagctcttcaagcagaccgccagctcttctaactgatggagccc	436
Db	140116	TCAICGTTCTTGGGCCAGCTCTTCAGCCAGCACCCCGACGGTCCTTACTCTATGAGCCC	140057
QY	437	gcgtggcagtgtgaacacacctgtcgcaggcgagcgcgcgaacgctgcacatggccgtg	496

CGTGGCATGTGTGGACCAACCTGTGCAGGCGCGGCAACGGTGCACATGGCCGIG 9107

Quality coverage: 7.39 in Q20 bases; pulse field per contig estimation.
Quality coverage: 4.59 in Q20 bases; sum-of-contigs estimation.


```
QY 497 cccgacctgatcgctctatctttttgtgagatggagctgtttgatgcctacatgcc 556
Db 61679 CCGGACCTGANGCGCTTAICTTTTGTGCGACATGGAGCTGTTCATGCTCAATGCCA 61738
QY 557 cagagccgaacactgtcgctcttttcaacttgggcaacagcgcgcgctgtgctcgcg 616
Db 61739 CAGAGCCGAACACTGTGCGCTTTTCAACTGGGCAACGAGCGCGGCTGTGCTCGCG 61798
QY 617 cccgctgacgagcctttcccgagggcaccatcagcaagcaggaagctatgaagacactg 676
Db 61799 CCGGCTGACGGCTTTCCCGAGGACACATCAGCAAGAGGAGGTATGCAAGACACTG 61858
QY 677 tgcacgagcagcaattcagcctgcccgggagggcctgcctctcagcacaagtggtg 736
Db 61859 TGCACGGGAGGACCATTCAGCTGCGCGCGGAGGCGCTGCCCTCTCAGCACCAGTGTG 61918
QY 737 ctaagaggtgtgctcttcaactcaggtgtctctaccgctgctcagcgaccccgcg 796
Db 61919 CTCAGGAGGTGCGCTTCTTCAACTGCAGGTGCTCTACCGGCTGCTCAGGACCCCGCG 61978
QY 797 ctcaactgtcgactgtgacctggtgcgagcccgcggtgtgtggcaccacagcgcaagtgg 916
Db 61979 CTCACCTGCGATCGTGTGCACTTGTGCGGACCCGCGGCGGCTGTGCGCTCCCGGAG 62038
QY 857 gcgcgggcgactgtgacacgacgacacgacgacgacgacgacgacgacgacgacgacg 916
Db 62039 GCGGCGGCGGCGGACGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 62098
QY 917 gtggagcgacacacacacacacacacacacacacacacacacacacacacacacacac 976
Db 62099 GTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 62158
QY 977 ggcgagcgcgacacacacacacacacacacacacacacacacacacacacacacacacac 1036
Db 62159 GCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62218
QY 1037 ttcgagagcctgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1096
Db 62219 TTCGAGGAGCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62278
QY 1097 ctgacccctcacccacacacacacacacacacacacacacacacacacacacacacacacac 1156
Db 62279 CTGACCCCTCAGCCACACGCTGAGGCGGCTGGATCCACACATCACCACGCGGTGCGGATC 62338
QY 1157 ggaagcaactagagccttcacactctgtctaggaatggcgcaacgctctccagggc 1216
Db 62339 GGAAGCAACTAGAGCGGCTTCCATCTCGTCTAGGATGCGCGCAACGCTCTCCAGGCGC 62398
QY 1217 tggcgcaacgcttgccctcactaagatcctgcgctgscggaggtgtgcgcccgcgcg 1276
Db 62399 TGGCGCACGCTTGCCTTTCACTAAGATCTGCGCGGTGCGAGGAGTGTGCGCGCGCGCG 62458
QY 1277 ctgcagctgtgggtctacccgctgtgtactctgcgacacagcgctgaactcacactg 1336
Db 62459 CTGCAGCTGTGGGTCTACCGGCTGTGTACTCTCGGACACGAGCGGTGACCTCACCCTG 62518
QY 1337 gatctgtgtgcacagggccagacacttcagctggcgatcgctgactgactgagaaactc 1396
Db 62519 GATCTGTGTGTGCCACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62578
QY 1397 tgggcttagagcaagcccgaaactgtgtgcagggcccgaggaagcgactgcatgtgtg 1456
Db 62579 TGGGCTTAGAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62638
QY 1457 aaagagctggggcgatgggaa-caggtccctactatcaacgggaggttggggtcc 1515
Db 62639 AGAGGAGGCTGGGCGCATGGGAGCAGGCTCCTACTATCAACCGGAGTTGGGGTCC 62698
QY 1516 t-ccctgaagtgaagcaagactgcagctttctctctctctctctctctctctctctctct 1574
Db 62699 TCCCTGAAGTAGGCAAGGACTGCAGGTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 62758
```

```
QY 1575 gagttcttgagctgccttctctcatcaggtgcactcttcatgcaaaagcaactcttgcgc 1634
Db 62759 GAGTCTTCTGGAGCTGCCCTTCTCATCAGGTGCACTCTTCTATGG-AAAGCAACTCTTGGCC 62817
QY 1635 ctactcttctggcgagggagtgtaagtactgtactgtactgtactgtactgtactgtactgt 1694
Db 62818 CTACTCTTCTGGGCGCAGGGAGTAAGTTACTGCTAAATTAATTAATTAATTAATTAATTAAT 62877
QY 1695 cgggtgcggtgctcactgtactgtactgtactgtactgtactgtactgtactgtactgtact 1754
Db 62878 CGGTTGCGGTGCTCATGCTCTAATPCCAGCAITTTGAGAGGCTGAGCGGCTGATCA 62937
QY 1755 cctgaggttcagga-ttcaaaacacagcctggcccaacatagtaaaacccctctctactaaa 1813
Db 62938 CCTGAGGTGAGGAGTTCGAAACACGCTGGCCACATAGTGAAACCCCTCTCTACTAAA 62997
QY 1814 aatgcacaaatagtcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1873
Db 62998 AATGCACAAATAGTCTCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 63057
QY 1874 gagtgggaaatcacttgagctcgaactcgaactcgaactcgaactcgaactcgaactcgaact 1933
Db 63058 GAGTGGGAGGAGTACATGAGCTCCAGAGTGGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 63117
QY 1934 ctgcacccctagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1993
Db 63118 CTGCACCCCTAGCTTGGGTGGCGAGCAAGACTCTATCAAAAAATAAGTAATAATTTGT 63177
QY 1994 tcaaaagtctgc 2006
Db 63178 TCAAAAGTCTGTC 63190

RESULT 5
AF219991
LOCUS
DEFINITION
AF219991 71503 bp DNA PRI 26-OCT-2000
Homo sapiens intestinal N-acetylglucosamine-6-O-sulfo-transferase
(CHST5) and corneal N-acetylglucosamine-6-O-sulfo-transferase
(CHST6) genes, complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
AF219991.1 GI:11023147
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 71503)
AUTHORS
Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T.,
Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
Macular corneal dystrophy type I and type II are caused by distinct
mutations in a new sulphotransferase gene
Nat. Genet. 26 (2), 237-241 (2000)
JOURNAL
MEDLINE
PUBMED
11017086
REFERENCE
2 (bases 1 to 71503)
AUTHORS
Akama,T.O. and Fukuda,M.N.
TITLE
Direct Submission
JOURNAL
Submitted (29-DEC-1999) Glycobiology Program, The Burnham
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1. .71503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
/clone="CITB-483K2"
358. .461
/rpt_family="L2"
repeat_region /rpt_type="dispersed"
complement(462..771)
/rpt_family="Alu"
/rpt_type="dispersed"
```

Fri Feb 1 20:20:49 2002

repeat_region	772. .945 /rpt_family="L2" /rpt_type-dispersed	repeat_region	/rpt_type-dispersed 15668. .15718 /rpt_family="Alu"
repeat_region	1169. .1338 /rpt_family="MIR"	repeat_region	15977. .16098 /rpt_type-dispersed /rpt_family="MIR"
repeat_region	/rpt_type-dispersed complement(1553. .1804) /rpt_family="Alu"	repeat_region	complement(17298. .17433) /rpt_type-dispersed /rpt_family="MIR"
repeat_region	2616. .2916 /rpt_family="Alu"	repeat_region	complement(18001. .18310) /rpt_type-dispersed /rpt_family="Alu"
misc_feature	3309. .3532 /note="similar to EST H12297 (EST cluster Hs.31147)"	repeat_region	complement(18420. .18591) /rpt_type-dispersed /rpt_family="MIR"
misc_feature	3865. .4028 /note="similar to EST H12297 (EST cluster Hs.31147)"	repeat_region	18732. .18767 /rpt_type-dispersed /rpt_family="Simple_repeat"
repeat_region	4099. .4406 /rpt_family="Alu"	CDS	18967. .20202 /rpt_type-tandem
repeat_region	4793. .5250 /rpt_family="L2"		/gene="CHST5" /note="I-glcNAc6ST; carbohydrate sulfotransferase 5; expressed in the small intestine and colon"
repeat_region	/rpt_type-dispersed complement(5251. .5329) /rpt_family="MIR"		/codon_start=1 /product="intestinal N-acetylglucosamine-6-O-sulfotransferase"
repeat_region	/rpt_type-dispersed complement(5768. .6080) /rpt_family="Alu"		/protein_id="AAG26326.1" /db_xref="GI:11023148"
repeat_region	/rpt_type-dispersed complement(6081. .6265) /rpt_family="Alu"		/translation="MGMRVPPKVAHSTRPPAARMWLPRESSKTVTVLLAQTCLL LFIIISRPSPAGGSDVHVLVSSWRSGSFLGQSFSDHDFYLMPEAWHVTTLL SQGSAAILHMAVRDLMRSIFELCDMDVDPAYMQSNLSAFFNWTALRCPACSAF PRGTISKODVCKTCTROPFSLAREACRSYSHVLYKEVFNLOWLYVLLSDPALNLR IVHLVROPRAVLRSKRAAGPILARDNGIVLGTNGKVEADPHLRIRVWCRSHVRIAE AATLPPPTFLGRVRLVDFEDLAREPLAEIRALYFTGLITLTPOLEATHNTHGSGI GKPEAFHTSSRNARNVSAWRHALPFTKLKRVQEVCAQALQLLGYRPVYSADQQRDL TLDIVLPRGPDHFSWASPD"
misc_feature	6111. .6567 /note="similar to EST AA001593 (EST cluster Hs.6853)"		20504. .20806 /rpt_family="Alu"
repeat_region	8035. .8350 /rpt_family="Alu"	repeat_region	/rpt_type-dispersed 20869. .20961 /rpt_family="L1"
misc_feature	10545. .11049 /note="similar to EST AI950923 (EST cluster Hs.6853)"	repeat_region	20962. .21200 /rpt_family="MER4-group"
repeat_region	11043. .11195 /rpt_family="Alu"	repeat_region	21325. .21429 /rpt_type-dispersed /rpt_family="MER4-group"
repeat_region	/rpt_type-dispersed 11200. .11222 /rpt_family="Simple_repeat"	repeat_region	21430. .21722 /rpt_type-dispersed /rpt_family="Alu"
repeat_region	11570. .11871 /rpt_family="Alu"	repeat_region	21725. .22007 /rpt_type-dispersed /rpt_family="Alu"
repeat_region	/rpt_type-dispersed complement(12065. .12198) /rpt_family="L1"	repeat_region	22008. .22040 /rpt_type-dispersed /rpt_family="Simple_repeat"
repeat_region	/rpt_type-dispersed complement(12205. .12493) /rpt_family="Alu"	repeat_region	22041. .22313 /rpt_type-tandem /rpt_family="MER4-group"
repeat_region	/rpt_type-dispersed complement(12813. .13110) /rpt_family="Alu"	repeat_region	22325. .22593 /rpt_type-dispersed /rpt_family="Alu"
mRNA	join(14185. .14228,16169. .16263,17712. .20821) /gene="CHST5" /product="intestinal N-acetylglucosamine 6-O-sulfotransferase"	repeat_region	22594. .22949 /rpt_type-dispersed /rpt_family="MER4-group"
gene	14185. .20821 /gene="CHST5"	repeat_region	23116. .23221 /rpt_type-dispersed /rpt_family="Simple_repeat"
repeat_region	14681. .14713 /rpt_family="Simple_repeat"	repeat_region	23344. .23400 /rpt_type-tandem /rpt_family="MaLR"
repeat_region	/rpt_type-tandem complement(14715. .14990) /rpt_family="Alu"	repeat_region	23433. .23456 /rpt_type-dispersed
repeat_region	/rpt_type-dispersed complement(14995. .15288) /rpt_family="Alu"		
repeat_region	/rpt_type-dispersed complement(15292. .15589) /rpt_family="Alu"		

Fri Feb 1 20:20:49 2002

us-09-593-828-3.rge

```

6
RESULT
AC009163
LOCUS
DEFINITION
AC009163 Homo sapiens chromosome 16 clone RP11-77K12, WORKING DRAFT
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 436185
Center clone name: RPCI-11-77K12
-----
Summary Statistics
Consensus quality: 132269 bases at least Q40
Consensus quality: 157541 bases at least Q30
Consensus quality: 162978 bases at least Q20
Estimated insert size: 63060; agarose-fp estimation
Estimated coverage: 1.4 in Q20 bases; sum-of-contigs estimation
Quality coverage: 3.87 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 51 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1020: contig of 1020 bp in length
* 1021 1120: gap of unknown length
* 1121 2488: contig of 1368 bp in length
* 2489 2588: gap of unknown length
* 2589 4067: contig of 1479 bp in length
* 4068 4167: gap of unknown length
* 4168 5211: contig of 1044 bp in length
* 5212 5311: gap of unknown length
* 5312 6416: contig of 1105 bp in length
* 6417 7887: gap of unknown length
* 7888 9088: contig of 1371 bp in length
* 9089 9189: contig of 1101 bp in length
* 9189 10211: gap of unknown length
* 10212 11713: contig of 1402 bp in length
* 11714 11813: gap of unknown length
* 11814 13073: contig of 1260 bp in length
* 13074 13173: gap of unknown length
* 13174 14854: contig of 1681 bp in length
* 14855 14955: gap of unknown length
* 14955 16909: contig of 1955 bp in length
* 16909 17010: gap of unknown length
* 17010 18668: contig of 1659 bp in length
* 18668 18768: gap of unknown length

```

```

18769 20273: contig of 1505 bp in length
20274 20373: gap of unknown length
20374 21426: contig of 1053 bp in length
21427 21526: gap of unknown length
21527 22910: contig of 1384 bp in length
22911 23010: gap of unknown length
23011 25171: contig of 2161 bp in length
25172 25271: gap of unknown length
25272 27706: contig of 2435 bp in length
27707 27806: gap of unknown length
27807 29083: contig of 1277 bp in length
29084 29183: gap of unknown length
29184 31233: contig of 2050 bp in length
31234 31333: gap of unknown length
31334 33854: contig of 2521 bp in length
33855 33954: gap of unknown length
33955 36107: contig of 2153 bp in length
36108 36207: gap of unknown length
36208 38156: contig of 1949 bp in length
38157 38256: gap of unknown length
38257 39887: contig of 1731 bp in length
39888 40087: gap of unknown length
40089 41609: contig of 1522 bp in length
41610 41709: gap of unknown length
41710 43400: contig of 1691 bp in length
43401 43500: gap of unknown length
43501 45245: contig of 1745 bp in length
45246 45345: gap of unknown length
45346 47679: contig of 2334 bp in length
47680 47779: gap of unknown length
47780 49623: contig of 1844 bp in length
49624 49723: gap of unknown length
49724 51988: contig of 2265 bp in length
51989 52088: gap of unknown length
52089 54400: contig of 2312 bp in length
54401 54500: gap of unknown length
54501 57337: contig of 2837 bp in length
57338 57437: gap of unknown length
57438 59990: contig of 2553 bp in length
59991 60090: gap of unknown length
60091 62554: contig of 2464 bp in length
62555 62554: gap of unknown length
62555 65581: contig of 2927 bp in length
65582 65681: gap of unknown length
65682 70061: contig of 4380 bp in length
70062 70161: gap of unknown length
70162 74922: contig of 4761 bp in length
74923 75022: gap of unknown length
75023 79848: contig of 4826 bp in length
79849 79948: gap of unknown length
79949 83052: contig of 3104 bp in length
83053 83152: gap of unknown length
83153 87068: contig of 3916 bp in length
87069 87168: gap of unknown length
87169 93512: contig of 6344 bp in length
93513 93612: gap of unknown length
93613 98370: contig of 4758 bp in length
98371 98470: gap of unknown length
98471 102251: contig of 3781 bp in length
102252 102351: gap of unknown length
102352 107787: contig of 5436 bp in length
107788 107887: gap of unknown length
107889 112454: contig of 4567 bp in length
112455 112554: gap of unknown length
112555 120360: contig of 7806 bp in length
120361 120460: gap of unknown length
120461 126854: contig of 6394 bp in length
126855 126954: gap of unknown length
126955 134018: contig of 7064 bp in length
134019 134118: gap of unknown length
134119 146223: contig of 12105 bp in length
146224 146323: gap of unknown length
146324 157038: contig of 10715 bp in length

```

FEATURES	source
* 157039	157138: gap of unknown length
* 157139	179065: contig of 21927 bp in length.
	Location/Qualifiers
	1..179065

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-77K12"
/cloned_lib="Rpci human BAC library 11"
47243 a 39973 c 38944 g 47214 t 5691 others
BASE COUNT
ORIGIN

```

```

Query Match      79.7%; Score 1628.8; DB 2; Length 179065;
Best Local Similarity 95.6%; Pred. No. 2e-260;
Matches 1729; Conservative 0; Mismatches 72; Indels 7; Gaps 5;

QY 201 ggcgccagcgcccgcatgtgtgctccacggtctccagcagacagatgacagtgtccc 260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166723 GGCCTCCAGCAGTTCAGCATGTGGCTTCGCGCGGTCTCCAGCACACGAGTGACGGCGTCC 166782
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 261 tcttgccacagaccactgcctctgtcttctatcatctccgggcagggccctatccc 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166783 TCTTGGCGGAG---ACCTTCCTCCCTCTTCTGGTTCCTGGGCGCAGGGCCCTCGTCTCC 166839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 321 cagcggcgcgagagatcgtgtgcagtgctgtgtctctctgtggtcgtcggtctcat 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166840 CACGAGCGCGGAGGCGCGCGTGCATGTGCTGGTGTCTCTCTGCTGGCGCTCGGCTCGT 166899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 381 ccttcttgggcgagctctctcagcagacaccccgactcttctaactgatggagcccgct 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166900 CCTTCGTGGGCCAACTCTTCACACGACACCCGACGTCTTCTACCTAATGGAGCCCGGT 166959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 441 ggcattgtggaccacccctgtcgcaggcgagcgcgcaagctgcacatggccgtgcgcg 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166960 GGCACGTGTGGACCACTCTGTCGAGGCGCGGCGCAACGCTCACATGCTGTGCGCG 167019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 501 acctgatgcgtctatatctttttgtgcacatggcagctgtttgatgccttaactgccacaga 560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 167020 ACCTGGTGGCTCGCTCTTCCTGTGGCAGATGGACGTTGTGATGCTATCTGCGCTTGGC 167079
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

QY	561	gcgaaacactgtccgccttttcaactggcaacagcgcgcgcctgctcgcgcgcgcg	620
Db	167080	GCGCAAACCTGTCCGACCTCTCCACGTGGCCGTGACCTGTGCACCTGTGCTCCGCCCGC	167139
QY	621	ctgcagcgcctttcccgaggcaaccatcgaacgaggacgtatgcaagacactgtgca	680
Db	167140	CTGCAGTGCCTTCCCGAGCGGCATCAGAAAGCAGACGTGTCTCAGACACTGTGC	167199
QY	681	cgcgacgacattcagctggcccgaggagcctgcgctcctacagccacgtggtgctca	740
Db	167200	CSCGCACGCAATCAGCTGGCCCGGAGCCTGCCGTCTCACGCCACGTGGTGCTCA	167259
QY	741	aggaggtgcgcttcttcaacctgcaggtgctctaccgcgtcctcagcagcccgagctca	800
Db	167260	AGGAGGTGCGCTTCTTCAACTGCAAGTGCCTACCCGCTGCTACCGCGACCTCCGCGCTCA	167319
QY	801	acctgcacatgctgcacctgctgpcgaccccgggccgctgctgcgtcccgggagcgcg	860
Db	167320	ACCTGCGCATCTGTGCACCTGTGTGGCGACCCGCGGCGCTGCTGGCTCCCGGGAGGCGG	167379
QY	861	cgggcccgatactggcacgcgcacacgcatcgtgctggccacacggcagtggggtgg	920
Db	167380	CSCGCCCGATACTGGCAGCGCACACAGCGATCTGTCTGGGACCAACGGCAAGTGGGTGG	167439
QY	921	aggcgacacctcaactgcgcctgattcgccagggtgaccgcagcacaatgcgcacgcgcg	980
Db	167440	AGSCCGACCTCACCCTGGGCTGTATTCGGAGGTGTGCCGAGGCACCTGGCGCATCGCCG	167499
QY	981	aggcgccacactaagccgcacaccttcttcgcggcgccgtaccgcctggtgcgcttcg	1040
Db	167500	AGSCCGGCACACTCAAGCCGCCACCCCTTCTTGCGCGGCCGCTACCGCTGGTGGCGCTTCG	167559

RESULT	7
AF176839	
LOCUS	
DEFINITION	

LOCUS	AF176839	1462 bp	DNA	PRI	22-SEP-1999
DEFINITION	Homo sapiens	intestinal	N-acetylglucosamine	6-O-sulfo	transferase

Fri Feb 1 20:20:49 2002

(I-GlcNAc-6-ST) gene, complete cds.

ACCESSION
AF176839
VERSION
AF176839.1
KEYWORDS
GI:5917707
SOURCE
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1462)
Lee, J.K., Bhakta, S., Rosen, S.D. and Hemmerich, S.

AUTHORS

Cloning and characterization of a mammalian
N-acetylglucosamine-6-sulfotransferase that is highly restricted to
intestinal tissue

JOURNAL

Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

MEDLINE

99423499

REFERENCE

2 (bases 1 to 1462)
Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.

AUTHORS

Direct Submission
Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA

JOURNAL

Location/Qualifiers

FEATURES

1..1462

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q23.1"

/tissue_type="intestine"

<229..>1401

/gene="I-GlcNAc-6-ST"

/product="N-acetylglucosamine 6-O-sulfotransferase"

<229..>1401

/gene="I-GlcNAc-6-ST"

229..1401

/gene="I-GlcNAc-6-ST"

/note="Possibly involved in biosynthesis of mucosal
ligands for L-selectin"

/codon_start=1

/product="N-acetylglucosamine 6-O-sulfotransferase"

/protein_id="AAD56001.1"

/db_xref="GI:5917708"

/translation="MWLPFSKIVTVLLAQTTCLLFIISRPSPAGGEDRVHV
LVLSWRSGSFLQFSDHPDVFYMEPAWHTLISQSAATLHMVARDLIMRIFL
CDMDVDFAYMFOQRNLISAFFWATSFALCSPPACSAFPRGTSKODVCKTLCTROPFS
LAREACRSYSHVVLKEVRFNLDVLLPLSDPALNLRVHLVEDPRAVLRSREARAPI
LARDNGVILGTNGKWEADPHLRIFREYCRSHVRAEALNKPPELGRVRLVRED
LAREPLAETRALYAFYTLTLPQLEAWTHNTHGSGIKPIEAFHTSSRNANVSOAW
RHAPFTKILRVOEVCAGALQLIGYRPVYSADQQRDLTLDLVLPRGPDHFSWASPD"

BASE COUNT 239 a 510 c 446 g 267 t

ORIGIN

Query Match 61.1%; Score 1248.6; DB 9; Length 1462;
Best local Similarity 99.7%; Pred. No. 3.3e-197;
Matches 1251; Conservative 0; Mismatches 4; Indels 0; Caps 0;

QY 197 cgggtggcccccagccgcgcagtggtggtccacggtttccagcaagacagtgacagtg 256

Db 208 CGAAGGCCCCAGCGCCGCGATGTGGTCCACGGTCTCCACAGACAGTACAGATG 267

QY 257 ctctctgcacagacacacgctgctctctctctctctctctctctctctctctctca 316

Db 268 CTCCTCTGGCAGACAGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 327

QY 317 tccccagccgcccagagagtcgtgacacgctggtgctgctgctgctgctgctgctggc 376

Db 328 TCCCCAGCCGCGCGAGGATCGTGTGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 387

QY 377 tcattcttttggccagctcttaagccagcaccacacgctctctctctctctctctctgagcc 436

Db 388 TCATCTCTTGTGGCCAGCTCTTCAGCCAGCACCACCGACGCTCTTCTACTGATGAGGCC 447

QY 437 gctggcagctgtgtagcacctctcgcaggcagcgcgcgcgcgcgcgcgcgcgcgcgcgc 496

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Db 448 GCCTGCAGTGTGTGGACCACTCTGTCGAGGCGAGCGCGCAACGCTGCACATGSCCGTG 507

QY 497 cgcgacctgatgcgtct 556

Db 508 CGGACCTGATGCGCTCTATCTCTTTTGTGGACATGACGCTGTGTGATGCTACATGCCA 567

QY 557 cagagccgaacacgtctcgcct 616

Db 568 CAGAGCGGAAACCTGTCCGCTTTTCAACATGGGCAACGAGCCGCGCTGTGCTCGCGG 627

QY 617 cccgctgcagcgcct 676

Db 628 CCGGCTGTGACGCGCTTTTCCCGAGGACCATCAGCAAGAGGAGCTATGAAGACACTG 687

QY 677 tgcacgagcagcgcct 736

Db 688 TGCACGCGGACGACATTCAGCTGTGGCCGGGAGGCTCCGCTCTCAGACCACTGGTG 747

QY 737 ctcaagaggtgagct 796

Db 748 CTCAGGAGGTGCGCTTCTTCAACCTGAGGCTCTTACCCGCTCTCAGCGAACCCTCGG 807

QY 797 ctcaacctgcatctgctgacacgtgctgagcaccgcccgcgcgcgcgcgcgcgcgcgc 856

Db 808 CTCACCTGCGCATCTGTCACCTGCTGCGGACCGCGCGCGCTGCTGCGCTCCCGGAG 867

QY 857 gggcgggccgcgatactggcagcagcagcagcagcagcagcagcagcagcagcagcag 916

Db 868 GCGCGGCGCCGATACTGTCAGCGGACACGCGCTGCTGCTGCGGACCAACGCAAGTGG 927

QY 917 gtggagggccgacctcactcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 976

Db 928 GTGGAGGCGGACCTACCTGCGGCTGATTCGGAGAGGTGTGCGGACCGACGCTGCGCATC 987

QY 977 ggcgagcgcgcacacactcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1036

Db 988 GCGGAGCGCGCCACACTCAAGCGCGCCACCTCTCTCTGCGGCGCGCTACCGCTTGGTGGC 1047

QY 1037 ttgagagacctgcccgcgggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1096

Db 1048 TTGAGGACCTGGCGGGAGCGCTGGCAGAGATCGCGCACTCTACGCTTTCACCGCG 1107

QY 1097 ctacacctcagccacagctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1156

Db 1108 CTGACCTCAGCGCACAGCTGCGGCGCTGGATCCACAACTATCACCACGCGGTGCGGATC 1167

QY 1157 gcaagcacaatcagcgcctccactctctctctctctctctctctctctctctctctctct 1216

Db 1168 GGCNAGCCATCGAGGCTTCCATPACTTCTGTAGGAATGCGCAACGCTCTCCAGGCC 1227

QY 1217 tggcgccacgcttgccctcactcaatagactcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1276

Db 1228 TGGCGCCACGCGTGTGCGCTTCTACTAAGATCTGCGGCTGCGGAGGAGTGTGCGCGCGCG 1287

QY 1277 ctgcagctgctggctacccgctctgctctctctctctctctctctctctctctctctct 1336

Db 1288 CTGAGCTGTCTGGGCTACCGGCTCTGTACTCTGGGACCAAGCAGCGTGACTACCTACCTG 1347

QY 1337 gatctggtgctgcaacgagggcccgagccactctgctgctgctgctgctgctgctgctg 1396

Db 1348 CATCTGGTGTCTGCCACGAGGCCAGACCATCTCAGCTGGGATCGCTCTGACTGAACTC 1407

QY 1397 tggccttagagcagcagcccccgaactgtgtgcagggcccgagggagcgactgcat 1451

Db 1408 TGGGCTTTAGAGCAGCGCCCGAATGTGTGTGCCAGGCCCGAGAGCGGACCTGAT 1462

RESULT 8

AF280086

LOCUS

DEFINITION

ACCESION

3786 bp

mRNA

Homo sapiens N-acetylglucosamine 6-O-sulfotransferase

mRNA, complete cds.

PRI

20-FEB-2001

```

VERSION      AF280086.1  GI:12060803
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 3786)
AUTHORS      Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
             Rosen,S.D.
TITLE        Chromosomal localization and genomic organization for the
             galactose/ N-acetylgalactosamine/N-acetylglucosamine
             6-O-sulfotransferase gene family
JOURNAL      Glycobiology 11 (1), 75-87 (2001)
MEDLINE
REFERENCE    2 (bases 1 to 3786)
AUTHORS      Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and
             Rosen,S.D.
TITLE        Direct Submission
JOURNAL      Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
             3401 Hillview Ave., Palo Alto, CA 94304, USA.
FEATURES
             source
             location/Qualifiers
             1..3786
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="16"
             /map="16q23.1"
             1..3786
             /gene="GST4beta"
             /note="clusters with the intestinal N-acetylglucosamine
             6-O-sulfotransferase (GST4alpha) gene in GenBank Accession
             Numbers AF176838 and AF176839"
             189..1376
             /gene="GST4beta"
             /codon_start=1
             /product="N-acetylglucosamine 6-O-sulfotransferase
             GST-4beta"
             /protein_id="AAG48244.1"
             /db_xref="GI:12060804"
             /translation="MPLPRYSSTAVTALLLAQTLLFLVSRPSPSGAGEARVHVL
             DMSWRSGSEFVGLQHPDVFYLMPEPAHWITLSGSAATLHMAVRDLVRVFLC
             DMVDYDALPRLNLSLDFQWANSRALCSPACSAFPERGATSEAVCKPLCARQSFIL
             ARACRSYSHVLLKEVRFNQLVLYPLLSDFALNLRIVHLVDPRAVLRSEONAKAL
             ARDNGIVIGNGTWADPGLRVVREVRSHVRIAEAAITKPPDLRLGRVLRFEDEI
             AREPLAIRALYAFITGLSLTLPQLEAWIHNITHGSGPARREAPKTSRNALNVSQAR
             HALPFAKIRRVQELCALQALLGYRPVYSEDEQRNALDLVLPRLNGFTWASSTASH
             PRN"
             3106..3111
             /gene="GST4beta"
             /note="alternative signal"
             3734..3739
             /gene="GST4beta"
BASE COUNT   848 a 1030 c 1104 g 804 t
ORIGIN
Query Match      53.9%; Score 1101; DB 9; Length 3786;
Best Local Similarity 85.4%; Pred. No. 6.2e-173;
Matches 1314; Conservative 0; Mismatches 205; Indels 20; Gaps 7;
QY 140 tcagacatcccatgattgatcagcgttgagctctccagcagccctggacgg 199
DB 111 TCTAATCATCTCCATGGATTGTGATCAGCGTTGACGCTCTCCAGCAGCCCTGGACAS 170
QY 200 tggccccagcccgcatgtggtgccacggttctccagcaagacagtgcagtgctc 259
DB 171 TGGCCCCAGCAGCAGTACGATGTGGCTTCCCGCGCTCTCCAGCAGCAGCAGTGCAGCGCTC 230
QY 260 ctctggacagacacacgtgctctgtcttctatctatctcccgccagggccctcatcc 319
DB 231 CTCTGGCGCAG---ACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 287
QY 320 ccagccgcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 379

```

```

DB 288 CCAGCAGCGCGGAGCGCGCGCTGCATGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 347
QY 380 tctctcttggccagcgtcttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 439
DB 348 TCCCTTCGTGGGCCAACTCTTCAACAGCAGCCCGCAGCGCTTCTTACCTAATGAGCCGCG 407
QY 440 tggcatgtgtgacacacccctgctcgagggcagcgcgcaacgctcacatggcagcagcagcagc 499
DB 408 TGGCAGCGTGTGGACACACCTGTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 467
QY 500 gacctgatgcctctatcttcttggacatggacgtgtttgatgcctacatgccacag 559
DB 468 GACCTGTGCGCTCGCTCTTCTCTGTGGACATGGACGFTTTGATGCTATCTGCTTGG 527
QY 560 agcgaaacctgtccgctcttccaactgggcaacgagcgcgctgtgtctccgccc 619
DB 528 CGCGCGCAACCTGTCCGACCTCTTCCAGTGGCGCGTGGCGCGCGCGCGCGCGCGCGCGCG 587
QY 620 gctgcagcgccttcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 679
DB 588 GCCTGCAGTGCCTTTCCTCGAGGCGCATCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 647
QY 680 acgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 739
DB 648 GCCTGCAGTGCCTTTCCTCGAGGCGCATCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 707
QY 740 aagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 799
DB 708 AAGCAGCGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767
QY 800 aacctgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 859
DB 768 AACTTACGCATCGTGCACCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 827
QY 860 gggggcccgatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 919
DB 828 ACAGCCAGAGGCTGTGCGCGCGTGCACACGCGCTGCTGGCGCGCGCGCGCGCGCGCGCGCG 887
QY 920 gaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 979
DB 888 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 947
QY 980 gaggcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1039
DB 948 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1007
QY 1040 gaggacctggcgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1099
DB 1008 GAGGACCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1067
QY 1100 acctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1159
DB 1068 AGTCTACGCCACAGCTCGAGGCGCTGGATCCATACATCACCCACGGATCTGGACCTGGT 1127
QY 1160 aagcgaatcaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1219
DB 1128 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1187
QY 1220 cgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1279
DB 1188 CGCATATCGCTGCCCTTTGGCCAGATCCGCGCGTGCAGGAACTGTGCGTGTGGTGGCGCTG 1247
QY 1280 cagctgtgggtacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1339
DB 1248 CAGCTGTGGGTACCGGCTGTGTACTCTGAGGACGAGGAGCGCGCGCGCGCGCGCGCGCGCG 1307
QY 1340 ctggtgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1397
DB 1308 CTGCTGTGCCACGAGGCGCTGACGCGCTTCACTTGGGATCATCCACCGCGCTCGACCC 1367
QY 1398 gggccttagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1455
DB 1368 CGAAATATTGGAGGCGCACAGTTGTAGAGGCGCTAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1427

```

Fri Feb 1 20:20:49 2002

QY 1456 gaaaaggaactggcgacatgggaa-caggtccctactatacaacagcgaggttggtg- 1513
 Db 1428 CAGAGGGGCTGGGGCGCAGGAGAGCAGGTCCTATATTGACCAAGGAGTTGTGTA 1487
 QY 1514 -----cctccctgaagtaagaagaaactgaactgaacttcttctcctgaatctcg 1564
 Db 1488 CGACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1547
 QY 1565 gttttcttgatcttcttgagactgcttctctcatcaggtgcactctctcatgaaagca 1624
 Db 1548 GTTTTCTTGTAGTCTCTGAGACTGCTTCTCATCAGGTGACTCTTCATGG-AAAGCA 1606
 QY 1625 actctgcccctactcttcttgagactgcttctctcatcaggtgcactctctcatgaaagca 1663
 Db 1607 ACTCTTGGCCCT-CTCTCTCTGAGCAGAGGTGTGCGTT 1644

RESULT 9
 AF219990 2544 bp mRNA PRI 26-OCT-2000
 LOCUS Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase
 DEFINITION (CHS76) mRNA, complete cds.
 ACCESSION AF219990
 VERSION AF219990.1 GI:11023145
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2544)
 Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T.,
 Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
 Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
 Macular corneal dystrophy type I and type II are caused by distinct
 mutations in a new sulfotransferase gene
 Nat. Genet. 26 (2), 237-241 (2000)
 JOURNAL Nat. Genet. 26 (2), 237-241 (2000)
 MEDLINE 20472330
 PUBMED 11017086
 REFERENCE 2 (bases 1 to 2544)
 AUTHORS Akama,T.O. and Fukuda,M.N.
 TITLE Direct Submission
 JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham
 Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES Location/Qualifiers
 1..2544
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="16q22"
 1..2544
 /gene="CHS76"
 /note="responsible for macular corneal dystrophy"
 693..1880
 /gene="CHS76"
 /function="transfers sulfate to 6-OH of GlcNAc in keratan
 sulfate"
 /note="C-GlcNAc6ST"
 /codon_start=1
 /product="corneal
 N-acetylglucosamine-6-O-sulfotransferase"
 /protein_id="AAG26325.1"
 /translation="MWIPRVSTAVTALLAQFLFLFLVRSRGPSPAGGEARHVHL
 VLSWRSGSFGVQLNPDVFLMEPAHWHTILSQSAATLHMAYRDLVRSVFLC
 DMEDFAYLPPRRNLSDLFQWVRAICSPACSAFPGAISSAEVCKPLCARQSTFL
 AREACRSYSHVYLKVFYRFTNLQVLYPLISDPALNRLVHLPVPPVLRVRSRGTAKAL
 ARDNGLVIGTNGVADPCLRVRCVSHVRIAPAAIILKPPPLRGVRLVRFDEL
 AREPLAEIRALYAFGLSLTPQLPDAWHTNTHSGSGPGRAREAEKTSRNALNVQAWK
 HALPFAKIRRVQELCAGALQLLRPVYSEDEQNALDLVLPRLNGFTWASSTASH
 PRN"
 460 a 799 c 733 g 552 t

Query Match 51.5%; Score 1052; DB 9; Length 2544;
 Best Local Similarity 85.0%; Pred. No. 8.3e-165; Indels 18; Gaps 5;
 Matches 1240; Conservative 0; Mismatches 200;
 QY 140 tggagacatctccatgatttgatcagcgttgagctctccacagcagccctggagcg 199
 Db 615 TCTAATCATCTCCATGATTTGTGATCAGGTTGACGCTCTCCACAGCAGTACCGCGCTC 674
 QY 200 tggcccccagccgcccagtcggtccacggttctccagcaagacagtgacagtgc 259
 Db 675 TGGCCCCCAGAGTACAGTGTGCTGCGCGCTCTCCACAGCAGTACCGCGCTC 734
 QY 260 ctcttgccacagacacactgctctctctctctctctctctctctctctctctctct 319
 Db 735 CTCTTGGCGCAG---ACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 791
 QY 320 ccagccggcgcgagagatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg 379
 Db 792 CCAGCAGCGCGGAGCGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 851
 QY 380 tctcttggccagctctctcagccagcaccacacacacacacacacacacacacacac 439
 Db 852 TCCITCGTGGGCCAACTCTCAACCCAGCAGCCCGACGCTTCTACCTAATGAGCGCGG 911
 QY 440 tggcattgtggacacacactgctcagggcagcgcgcaacactgcacatggcctggcg 499
 Db 912 TGGCAGCTGTGGACACACCTGTGCGAGGGCAGCGCGCGCAACGCTGCACATGGCTGTG 971
 QY 500 gacctgatcgct 559
 Db 972 GACCTTGGTGGCT 1031
 QY 560 agccgaaacactgctccgacctctctcactgggcaacagcgcgcgctgctcgcgcgc 619
 Db 1032 CGCCGCCAACCTGTCCGACCTCTCCAGTGCGCGCTGAGCGCTGCTGCTGCTGCTGCC 1091
 QY 620 gctgcagcgctctctcccgagcgacacacacacacacacacacacacacacacacac 679
 Db 1092 GCCTGAGTGCCTTCTCCGAGGCGCATCAGCAGCGAGCGCGCTGTGCAAGCCACATGTC 1151
 QY 680 acgcgagcaccattcagcctgcccggagcgctgcccctcctacacacacacacacac 739
 Db 1152 GCGCGCAGTCTCTACCTGCGCGCGGAGGCTCTGCGCTCTTACACCCAGCTGCTGCTC 1211
 QY 740 aaggagtgctcttctcactgacgtgctctcactcactgctcactcactcactcactc 799
 Db 1212 AAGGAGTGGCTCTCAACCTGCGAGTGTCTTACCCGCTGCTCAGCGACCGCGCTC 1271
 QY 800 aactgcgcactcgtgcacctggcgagaccccgcgcgcgctgctcctccggagcg 859
 Db 1272 AACCTACGATCGTGTGACCTGGTGGCGGACCGCGCGCGCTGTGCGCTTCCCGGAGCAG 1331
 QY 860 gggggcccgatactgacacagcaacgcatcgctggcgacacacacacacacacacac 919
 Db 1332 ACAGCCAAAGGCTCTGGCGCGTGTGACACGGCATCTGCTGTGGGACCAACAGCGCAC 1391
 QY 920 gaggcccgacctcactcactgctgctgctgctgctgctgctgctgctgctgctgct 979
 Db 1392 GAGCGCGACCCCGGCTTGGCGTGGTGGCGAGGTTGTGCGGTAGCGACGTACGATGCG 1451
 QY 980 gaggcccgccactcactcactcactcactcactcactcactcactcactcactcactc 1039
 Db 1452 GAGCGCGCACACTCAACCGCGCACCTTCTTGGCGCGCGCGCTTACCGCTGTGGCTC 1511
 QY 1040 gaggacctggcgaggacccctcctcctcctcctcctcctcctcctcctcctcctc 1099
 Db 1512 GAGGACCTGGCGGAGCGCGCTGGCGAATAATCCCTGCGCTCTACGCTTCTACTGGGCTC 1571
 QY 1100 accctcagcagcactcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1159
 Db 1572 ACTCTACGCCACACGCTCGAGGCGCTGGATCCATACATCACCGGATCTGACCTGTG 1631

QY	1566	ttttcttttagtctctggagctgcctctcatcgcagtgacacttctcatggaagaa	1625
Db	136379	TTTTCTTTGAGTCTCTGAGAGCTGCTTCTCATCAGGTGACATCTTCTATG	136321
QY	1626	ctttgccccctacattctcttctggcgagcgagggagtaatt	1663
Db	136320	CTCTGCCCCCT-CCTCCTCTGGCACAGGCTGCGGTT	136284
RESULT	11		
AC009105	208185 bp	DNA	HTG
LOCUS	208185 bp	25-APR-2001	
DEFINITION	Human sapiens chromosome 16 clone RP11-455E15, WORKING DRAFT SEQUENCE, 17 unordered pieces.		
ACCESSION	AC009105		
VERSION	AC009105.7	GI:13786304	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 208185)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Sequencing of Human Chromosome 16 Unpublished		
REFERENCE	2 (bases 1 to 208185)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Apr 25, 2001 this sequence version replaced gi:7689928. -----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: http://www.jgi.doe.gov ----- Project Information Center Project Name: 0 Center Clone Name: RPECI-11_455E15 ----- Summary Statistics Consensus quality: 180040 bases at least Q40 Consensus quality: 192095 bases at least Q30 Consensus quality: 195726 bases at least Q20 Estimated insert size: 195650; agarose-fp estimation Estimated insert size: 206585; sum-of-contrigs estimation Quality coverage: 6.57 in Q20 bases; agarose-fp estimation Quality coverage: 6.23 in Q20 bases; sum-of-contrigs estimation. * NOTE: This is a 'working draft' sequence. It currently * consists of 17 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 1062: contig of 1062 bp in length * 1162: gap of unknown length * 1163 2359: contig of 1197 bp in length * 2360 2459: gap of unknown length * 2460 3653: contig of 1194 bp in length * 3654 3753: gap of unknown length * 3754 5640: contig of 1887 bp in length * 5641 5740: gap of unknown length * 5741 7370: contig of 1630 bp in length * 7371 7470: gap of unknown length * 7471 8584: contig of 1114 bp in length * 8585 10014: contig of 1330 bp in length * 10015 10114: gap of unknown length * 10115 12451: contig of 2337 bp in length * 12452 12552: gap of unknown length * 12552 28715: contig of 16715 bp in length		

FEATURES	source
28716	28815: gap of unknown length
28816	43140: contig of 14325 bp in length
43141	43240: gap of unknown length
43241	53865: contig of 10625 bp in length
53866	53965: gap of unknown length
53966	72091: contig of 18126 bp in length
72092	72191: gap of unknown length
72192	87355: contig of 15164 bp in length
87356	87455: gap of unknown length
87456	109461: contig of 22006 bp in length
109462	109561: gap of unknown length
109562	131379: contig of 21818 bp in length
131380	131479: gap of unknown length
131480	157556: contig of 26077 bp in length
157557	157657: gap of unknown length
157657	208185: contig of 50529 bp in length.
location/Qualifiers	
1..208185	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="16"	
/clone="RP11-455B15"	
/clone_lib="RPC1 human BAC library 11"	
BASE COUNT	58183 a 45373 c 45592 g 57204 t 1833 others
ORIGIN	
Query Match	
Best_Local_Similarity	51.2%; Score 1046.4; DB 2; Length 208185;
Matches 1257; Conservative	0; Mismatches 201; Indels 20; Gaps
QY 201	ggccccagccgcgcgatgtggctgcacaggtttctccacagacagtgacagtgtccc 260
Db 107478	GGCCCCCAGCAGTACGATGTGGCTGCGCGCGTCTCCACACAGCAGTACCGCGCTCC 107537
QY 261	tcttgccacagaccacotgcctctgtcttctatctatctcccgccagggccctatccc 320
Db 107538	TCTTGGGCGCAG--ACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 107594
QY 321	cagcggcgaggagtgctgtgcagctgtgtgtctctctcgtggcgtcggactcat 380
Db 107595	CAGCAGCGCGGAGGCGCGCTGCATGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 107654
QY 381	ctctcttgccagctcttcagcagcaccccgacgtctctctatctatgtggagccggt 440
Db 107655	CCTTCGTGGGCCAACTCTTCAACACGACACCCGACGCTCTTACCTAATGGAGCCGCG 107714
QY 441	ggcatgtgtgacacacctctgcagggcagcgcgcaacgctgcacatggcgtgcgcg 500
Db 107715	GGCAGGTTGGACACCCCTGTCCAGGGCAGCGCCGCAACGCTGCACATGTGTGTGCGG 107774
QY 501	actgatgcgtctatcttttggcgcacatggacgtgtttgatgcctacatgccacaga 560
Db 107775	ACCTGTGCGCTCCGCTCTCTCTGTGGCAIAGGACGTGTTGATGCCATCTTGCCTTGGC 107834
QY 561	gccgaacctgtcgcgtcttttaactgggaacagcgcgcgtgtgtcgtgcgcgcgcg 620
Db 107835	GC CGCAACCTGTCCGACCTCTTCCAGTGGCGCGTGCAGCGTGCATGTGTCTGCCACCG 107894
QY 621	ctctcagcgccttccccagcgcacatcagcaacgagacgtatgcgaagacactgtgca 680
Db 107895	CCTGCAGTGCCTTTCGCCAGGCGCCATCAGCAGCAGGCGGTGTGCAAGCCACTGTGCG 107954
QY 681	cgcggcagccattcagcctggccgggagggcctcgcgcctcctaacagcagctggtctca 740
Db 107955	CGCGGCGAGTCTTACCTTGGCCCGGGAGGCCGTGCGGCTCTCAGACCACTGTGTGTCA 108014
QY 741	aggagtgcgctcttctaactgcaggtgtctataccgctgctcagcgaccccgcgctca 800
Db 108015	AGGAGGTGGGCTTCTTCAACCTTGCAGGTGCTCTACCCGCTCTCAGCGAACC CGCGTCA 108074
QY 801	acctgcgcactgtgcacctggtgcgcaccgcggcgctgctgcgtcccgggagcgcg 860

TITLE Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue

PRASE COUNT	415 a	586 c	550 g	438 t
Query Match	36.8%;	Score 751.2;	DB 10;	Length 1989;
Best Local Similarity	70.3%;	Pred. No. 4e-115;		
Matches 1046;	Conservative 0;	Mismatches 413;	Indels 17;	Gaps 3
Qy	202	gcccacagccgcgcgcgatgtgtgtgccacggcttcacagcaagaacacatgcacgtgtcct	261	
Db	91	GGCCCTAGCCAGAGGTATGCGGCTACCCCGCTTCTCCAGCACATGTCATGTTTCGCTCCCT	150	
Qy	262	cctgtcacagaccactcctctgcctctctcatctcccgccagggccctcatcccc	321	
Db	151	GATGGTACAG---ACTGGCATCTCTGGTCTTCCTGGTCTCCCGGCAAGTGCCATGCTCCCC	207	
Qy	322	agccgcgcgcgagggatcgtgtgcacgtgtgctgtcctcgtggcgctcgggctcacc	381	
Db	208	ACACAGCCCTGGGGAGCGTGTGCACGTCTGTGTACTGCTCTGCTGGCTCGGGCTCGTC	267	
Qy	382	ctcttgggccagctcttcacgcagcacccgcagcttctcaactgattggagcccggtg	441	
Db	268	CTTGCTGGGCCAGCTCTTCAGCCAAACACCCGATGCTCTTCTACCTGATGGAGCCGGCTTG	327	
Qy	442	gaatgtgtgacacacccctgtcgcagggcagcgcgcaacgctgcacatggccgtgtgcga	501	
Db	328	GCACCTCTGGGATACGTTGTCTGCAGGGGAGTGCCTCCCGACATCCCATATGGCCGTCGTGA	387	
Qy	502	cctgatgcctctatcttttggcacatggacgtgttggatgcctacatgccacagag	561	
Db	388	CTGTATCCCGCTCAGTGTCCATATGCACATGGACGCTATTGTATGCTACTACCTGCCCTGGCG	447	
Qy	562	ccgaacactgtccgcctttttcaactgggcaacagcgcgcgcgcgtgtgtcgtccgcgcgcg	621	
Db	448	CCGCACATCTCGGATCTCTTCCAGTGGCGGCTGAGCCGCGCATTTGTCTACCTCCCGGT	507	
Qy	622	ctgcagcgcctttcccgcaggaccacatcagcaagcaggacgtatgcaagacactgtcac	681	

AF176841		APF176841	1740 bp	DNA	ROD	22-SEP-1999
LOCUS	Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAC-6-ST) gene, complete cds.					
DEFINITION						
ACCSSION	AF176841					
VERSION	AF176841.1		GI:5917711			
KEYWORDS	house mouse,					
SOURCE	Mus musculus					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;					
REFERENCE	1 (bases 1 to 1740)					
AUTHORS	Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.					
TITLE	N-acetylglucosamine-6-sulfotransferase from a mammalian intestinal tissue					
JOURNAL	Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)					
MEDLINE	99423499					
REFERENCE	2 (bases 1 to 1740)					
AUTHORS	Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.					
TITLE	Direct Submission					
JOURNAL	Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA					
FEATURES	Location/Qualifiers					
Source	1..1740					
	/organism="Mus musculus"					
	/strain="C57Bl/6"					
	/db_xref="taxon:10090"					
	/chromosome="8"					
	/map="8E1"					
mRNA	/tissue_type="intestine" <405..>.1592					
gene	/gene="I-GlcNAC-6-ST"					
	/product="N-acetylglucosamine 6-O-sulfotransferase"					
CDS	<405..>.1592					
	/gene="I-GlcNAC-6-ST"					
	405..1592					
	/gene="I-GlcNAC-6-ST"					
	/note="possibly involved in biosynthesis of mucosal ligands for L-selectin"					
	/codon_start=1					
	/product="N-acetylglucosamine 6-O-sulfotransferase"					
	/protein_id="X595603.1"					
	/db_xref="GI:5917712"					
	/translation="MRLPFRFSVTMLIMVOTGILVLVSROYPSPPAGLCGRHVHLLVSRSSTSSFGQLFSDHPDYFLMEPAHWDTPLSGSAPALEMAVRDLIRSVFLCDMDVDAYLDLPERNTSDLFEQAWSEALCSPPVCPEAPRGNISSEVECKPUCATPFGLAOECASSYSHVVLEKEVFENLOVLPDISPALNRIVHLRDPRVALRVREQTAKALARDNIGLVGTNGTWEADRPURKVNEYCRSHVRIAEALHKPPPFDLRVLYREDARDPTLIREDLFATFGTLPIQLTWHITHGSPGARREAFTSRDALSYSQARHYLPFAKRIRRVGGGAQLLGYRSHVSELRQDSLIDLLIPRGMSDFKWASSTEQAPS"					
BASE COUNT	365 a 513 c 491 g 371 t					
ORIGIN						
Query Match	36.7%; Score 750.4; DB 10; Length 1740;					
Best Local Similarity	77.7%; Pred. No. 5.6e-115;					
Matches 920; Conservative	0; Mismatches 261; Indels 3; Gaps 1;					
QY	200 tgccccccagcgcgccgatggatgcttcacagtgttgcttcacagcacagtaaacagtgtctc	259				
Db	387 TAGGCCCTAGCCAGAGGATGCGGCTACCCGGTTTCCTCCAGCACTGCATTGCTC 446					
QY	260 ctcttggaacagacacctgctctctgtttcatcttcccaggcgagccctcatcc 319					
Db	447 CTGATGGTAGAC---ACTGGCACTCTGTGCTCTCTTGCTGGGCTCCCGGCAAGTGCATCGCTCC 503					
QY	320 ccagccgcgcgcgcgagatcgttgtgaactgtggtgtgtctctgtaggcctcaggctca 379					
Db	504 CCAGCAGGCGCTTGGGAGCGGTGCACTGCTGTGTAATGTCCTCTGTCGGCCTCGGGCTCG 563					
QY	380 tccttcttgccagctcttcagcagcaccocagcctcttatcattgatggagccgcq 439					


```

/note="assembly_fragment"
32754..39334
/note="assembly_fragment"
39435..46696
/note="assembly_fragment"
46797..55655
/note="assembly_fragment"
55756..62443
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
62544..72549
/note="assembly_fragment"
72650..83039
/note="assembly_fragment"
83140..92278
/note="assembly_fragment"
92379..107881
/note="assembly_fragment"
107982..126941
/note="assembly_fragment"
127042..146180
/note="assembly_fragment"
146281..167117
/note="assembly_fragment"
167218..186468
/note="assembly_fragment"
186569..215647
/note="assembly_fragment"
BASE COUNT 57545 a 47913 c 48482 g 59400 t 2307 others
ORIGIN

Query Match 27.8%; Score 568; DB 2; Length 215647;
Best Local Similarity 81.0%; Pred. No. 2.6e-85;
Matches 755; Conservative 0; Mismatches 150; Indels 27; Gaps 7;

Qy 1104 tcaagccacagctcagcctggatcccaacatcacccacgggtcgggatcggaagc 1163
Db 84402 TCATGCCAGACTCGAGSCCTGGATCCCAACATCACCCATGGGTGAGGATCGGCTGGC 84343

Qy 1164 caatcgagccttcacatcctcgtctaggaatgcgcgaacgtctccacggcctggcgcc 1223
Db 84342 CACTTGAGGCTCTAGACTTGTCTAGGAAGAGCTCAACGCTCTCCAGGCCAGCGCC 84283

Qy 1224 acgcttgccctcactaagatc-----ctgcgctcgaggaagtggt 1265
Db 84282 ACGGCTGCGCTTCACCAAGGCTCTTACCACCTGCTGCGCTTTCGCTGCGAGGAGTGT 84223

Qy 1266 gcgcgcgcgcctcagctgctggctacccgctgtgtactctgcggaccagcagctg 1325
Db 84222 CGGCGGCGCAGCTGCAGCTGCAGAACTACCGGCCCGGTGTACTCTGAGGACCGAGCAGC 84163

Qy 1326 acctcacctggatctggtctgcacagagccacagaccacttcagctggcctgcgctg 1385
Db 84162 ACCTCGGCCCTGGATCTGCTGCTGCACAGAGGCCCAACACCACTTCTAGTGTGCTTCTG 1385

Qy 1386 act--gagaactcggccttagaagaagcccgaaactgtggtcgccagcccgagaaagc 1443
Db 84102 ACTTCGGGAACCTGGGCTTCAGAGGAGGCTCAAGTGTGTGTCGCGAGTCCCTAGGAGGC 84043

Qy 1444 gactgatggtgaaagagctgggagcgaatgggaa--cagatccctactactcaacccg 1502
Db 84042 GACTGTACGGTGAGAGGAGCTGGGGTGCACGGGGAAGCAGGTCCTACTATACCAAG 83983

Qy 1503 gacttgggtctccctcgaagtaaggaagcactcaggtttctttctctctctattot 1562
Db 83982 GAGTTGGGTCTCTCCCTGAGCAGSAGAACTCGGCTTCCTTCTCTCTCTCTCTCTCT 83923

Qy 1563 cggtttctttgagctctcttgagctgccttctcagctcaggtgcactctcaggaag 1622
Db 83922 TCGTTTCTCTTTGAGTCACTGAGAGTGGCTTCTCATCAGGTGCTCACTCACTCACTGAGG 83864

```

```

Qy 1623 caactctggccctacctcttctggggcagcaggagtaagttaactgctaaattaa 1682
Db 83863 CAACTCTTGGCTTTACGTCCCTCTGGGCCAGGAGTAAGTT-CTGCTAAATTAATTA 83805

Qy 1683 tgtgtgccagccgggtgcgtgctcgtctgtaataccagcattttgagagctgag 1742
Db 83804 TGTGTCCAGCGCGGCGCGGTGGCTACGCTGTAATCCCAACACTTTGGGAGGCTGAG 83745

Qy 1743 gcgggtggtacacctgaggtcagga-ttcaaaaccagcctggccaacatagtgaacccc 1801
Db 83744 GTGGCCAGATCATTTGAAGTGGGAGTTCGAGAGCAGCCTGGCCAACATGGTGAACCT 83695

Qy 1802 ctctctactaaaatgcaaaattagtcgcggtggtggcacactctgttaatccagct 1861
Db 83684 GCCTCTAGTGAATAAGAAAATAGCTGGSCATGTTGGCGGGCGCTGTAGTCTAGCT 83625

Qy 1862 acttagagcctgaggtgggaaatacaacttgacctcaaaaggtggaggttgagct 1921
Db 83624 ACTTGGGAGGCTGAGGAGGAGAACTCACTTGAACCCAGGAGTGGAGGTTGCAGTGAGCC 83565

Qy 1922 gaaatcatgccactgca---ccctagcttgggtggcaaaagcaaaactcttcaaaaaat 1978
Db 83564 GAGACTGTGCCACTGCACTCCICAGCCCTGGGTGACAGAGCAAGACGCGCTCAAAAAA 83505

Qy 1979 aattaataaattgttcaaaagtcctgcgaa 2010
Db 83504 AAAAAAATAAATAAATTGTTCCAAA 83473

RESULT 15
AP002792/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-652A16 map 11q, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
ACCESSION
AP002792
VERSION
AP002792.1 GI:9188578
KEYWORDS
HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE
Homo sapiens DNA, clone:RP11-652A16.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 174187)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Homo sapiens 174,187 genomic DNA of 11q
Published Only in Database (2000) In press
2 (bases 1 to 174187)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.,
Direct Submission
Submitted (11-JUL-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-652A16
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159107 bases at least Q40
Consensus quality: 165997 bases at least Q30
Consensus quality: 169872 bases at least Q20
Insert size: 172187; sum-of-contigs
Quality coverage: 5.12x in Q20 bases; sum-of-contigs
-----

```

FEATURES	Location/Qualifiers
source	1. .174187

```

/obj_xref: 12000
/chromosome: "11"
/map: "llq"
/clone: "RP11-652A16"
1. 30947
/note="assembly_fragment"
31048. 49981
/note="assembly_fragment"
50082. 89634
/note="assembly_fragment"
69735. 89343
/note="assembly_fragment"
89444. 102389
/note="assembly_fragment"
102490. 114102
/note="assembly_fragment"
114203. 122892
/note="assembly_fragment"
122993. 130406
/note="assembly_fragment"
130507. 135937
/note="assembly_fragment"
136038. 140633
/note="assembly_fragment"
140734. 146349
/note="assembly_fragment"
148450. 152912
/note="assembly_fragment"
153013. 156057
/note="assembly_fragment"
156158. 159294
/note="assembly_fragment"
159395. 161770
/note="assembly_fragment"
161871. 164958
/note="assembly_fragment"
165059. 168161
/note="assembly_fragment"
168362. 170311
/note="assembly_fragment"
170412. 171211
/note="assembly_fragment"
171312. 172370
/note="assembly_fragment"
172471. 174187

```

```

misc_feature      /note="assembly_fragment"
136038..140633
/note="assembly_fragment"
misc_feature      140734..146349
/note="assembly_fragment"
148450..152912
/note="assembly_fragment"
153013..156057
/note="assembly_fragment"
156158..159294
/note="assembly_fragment"
159395..161770
/note="assembly_fragment"
161871..164958
/note="assembly_fragment"
165059..168161
/note="assembly_fragment"
168262..170311
/note="assembly_fragment"
170412..171211
misc_feature

```

BASE COUNT	50512 a	35755 c	36271 g	49647 t	2002 others
ORIGIN					

[illegible]

QY 1266 gcgcggcgcgctgcagctgctgggtaccggcctgtgtactctgcggaccagcagctg 1325
Db 136738 GCgcgcggcgcgctgcagctgcagaaCTAOCGGCCGFGTACTCTGAGGACGACGACG 136679

QY 1326 acctcaccctggatctggctgctccacgagccacaccacttcagctgggcatcgctg 1385
Db 136678 ACCTCGCCCTGGATCTGGTCTGTCTACAGGCGCAACACCACCTTCAGCTGGTTGTCTTCTG 136619

QY 1386 act--gagaactctgggccttagagcaagcccgaaactgtgtgcagccagcaggaagc 1443
Db 136618 ACTTCGGAACCTCTGGSCCTCAGAGAGGCCCTCAAGTTGTGGTCGCCAGTCTAGGAGGC 136559

QY 1444 gactgcatggtggaagagagctgggcgcacatgggaa--caggtccctactatcaaccgg 1502
Db 136558 GACTGTACGGTGGAGAGGAGTGGGCTGCACGGGGAAGCAGGTCCCTACTATTACCAG 136499

QY 1503 gqgtttggggtcctccctgaagtaagcagactgacaggttcacgtttcttctctctctct 1562
Db 136498 GAGTTGGGGTCTCTCCCTGAAGAGGCAAGAACTGCGGTTCTCTTCTCTCTCTCTCTCT 136439

QY 1563 oggtttctcttggatcttctgagctcttctgagctgcttctctcaggtgcactcttctcagaaag 1622
Db 136438 TCGTTTTCCTTTGAGTCATCTGAGGTGCTTCTCATCAGGTGCACTCAACATGG--GAAG 136380

QY 1623 caactcttgccctacctcttctgggcgagggagtaagtactgctaaattaaattaaa 1682
Db 136379 CAACCTTTGCCCTCTACGTCCTCTGGGCCCAGGGAGTAACTT-CTGCTAAATTAAATTANA 136321

QY 1683 tgtgtccagggccgggtgcgtgctcctgctgtaatcccaagcattttgagaggtgag 1742
Db 136320 TGTGTTCCAGGCCGGCGGCTGGGCTACGCTGTAAATCCCAACACTTTGGGAGGCTGAG 136261

QY 1743 gcgggtggtacacctgaggtcagga-ttcaaaaccagcctggccaacatagtgaaacccc 1801
Db 136260 GTGGCAGATCATCTGAAGTGGGAGTTCGAGAGCAGCTGGCCAAACATGGTGAACCCCT 136201

QY 1802 ctctctactaaaaatgcaaaaattagtcggtggtggcacactcctgtaatccagct 1861
Db 136200 GCCCTAGTGAATAACGAAAATTAGCTGGGCATGGTGGCGGCCCTGTAGTCTAGCT 136141

QY 1862 acttaggaggtgaggtgggaaatacttgactccaaaggtgaggttgagtaagct 1921
Db 136140 ACTTGGGAGGCTGAGAGGGAGAACTTGAACCCAGAGGTGGAGGTGCAGTGAGCC 136081

QY 1922 gaaatcagcactgcaccc 1941
Db 136080 GAGACTGTGCCTGCACTGCAC 136061

Search completed: January 31, 2002, 18:35:34
Job time: 12383 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 31, 2002, 18:15:53 ; Search time 315.85 Seconds
(without alignments)
5548.117 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggcctcgggtccactgtgct.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*

- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2044	100.0	2044	22	Human glycosyl sul
2	1973	96.5	2170	22	Human glycosyl sul
3	1827.2	89.4	2988	21	Human ORF1711
4	1753.6	85.8	160552	22	Human glycosyl sul
5	1111.4	54.4	1694	22	Human glycosyl sul
6	751.2	36.8	1989	22	Mouse glycosyl sul
7	400	19.6	2032	20	Human glycosyl sul
8	377	18.4	2065	21	Human glycosyl sul
9	354	17.3	1926	20	Human glycosyl sul
10	235.2	11.5	2156	19	Mouse glycosyl sul
11	211	10.3	1043	21	Glycosaminoglycan
					Human interleukin-

12	211	10.3	1601	21	AAF21313	Human low adenosin
13	211	10.3	1601	21	AAA35191	Human adenosine re
14	211	10.3	1601	21	AAZ46814	Interleukin-10 (IL
15	211	10.3	1618	12	AAQ10207	pH5C insert conta
16	211	10.3	1618	14	AAQ46958	Human cytokine syn
17	211	10.3	1618	21	AAZ86905	Human CSIF coding
18	211	10.3	1618	22	AAF90433	Human cytokine syn
19	211	10.3	1618	22	AAF82588	Human CSIF CDNA.
20	211	10.3	1618	22	AAF83186	Human CSIF protein
21	211	10.3	1645	21	AAZ58659	Human interleukin-
22	211	10.3	2797	22	AAH18167	Human interleukin-
23	211	10.3	15630	21	AAF21317	Human cDNA sequenc
24	211	10.3	15630	21	AAA35195	Human low adenosin
25	208.2	10.2	13865	19	AAV40401	Human adenosine re
26	208.2	10.2	13865	20	AAZ32165	Human tissue facto
27	207.4	10.1	3234	16	AAQ92781	Human cholesterol
28	205.8	10.1	700	22	AAH92085	Human thymopoietin
29	205.8	10.1	2932	13	AAQ25388	Human inflammatory
30	205.8	10.1	2932	20	AAZ32161	TXA2 receptor gene
31	205.8	10.1	2932	20	AAZ32162	Human thromboxane
32	205.6	10.1	15577	19	AAV35616	Human endothelial
33	205.6	10.1	32367	19	AAV35620	SHOX gene prelinin
34	205.4	10.0	48000	22	AAZ27996	Human SHOX (short
35	205.2	10.0	54548	21	AAZ45596	Human calcium sens
36	204.8	10.0	9456	22	AAI63326	DNA sequence of th
37	204.6	10.0	7240	22	AAH50570	Human kidney relat
38	204.6	10.0	7240	22	AAH31170	Human insulin rece
39	204.6	10.0	7240	22	AAH31267	Human insulin rece
40	204.6	10.0	7240	22	AAH31267	Human insulin rece
41	204.4	10.0	84607	20	AAZ90847	Human PACAP genomi
42	204.2	10.0	72928	20	AAZ18355	Human ASTH1J 5' ge
43	204.2	10.0	72928	21	AAA80253	Human ASTH1J 5' ge
44	204	10.0	119950	20	AAZ90201	Human yes1 gene.
45	203.4	10.0	1227	22	AAF45078	Human secreted pro

ALIGNMENTS

RESULT 1

AAAD02699

ID AAD02699 standard: cDNA; 2044 BP.

XX

AC AAD02699;

XX

DT 02-MAY-2001 (first entry)

XX

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenallitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocardiitis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1; ss.

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 218..1390

FT /*tag= a

FT /product= "Human glycosyl sulfotransferase-4alpha

FT (GST-4alpha)"

FT /note= "CDS is specifically claimed as SEQ ID NO: 4

FT in claim 6 (page no: 41) of the specification"

XX WO200106015-A1.

PN 25-JAN-2001.

XX


```

Db 1501 gggagttgggtctccctccctgaagtagcaaggagactcaggtttctctctcatt 1560
Qy 1561 ctcgtttctcttgagttctcttgagctgccttctctcaggtgcactctctcagaaa 1620
Db 1561 ctcgtttctcttgagttctcttgagctgccttctctcaggtgcactctctcagaaa 1620
Qy 1621 agcaactcttgccctacactctctcttgagcagggagtagtaactgctaaataaatta 1680
Db 1621 agcaactcttgccctacactctctcttgagcagggagtagtaactgctaaataaatta 1680
Qy 1681 aatgtgtccagcgcgggtgcgtctcagctgcctgttaatccagcattttgagagctg 1740
Db 1681 aatgtgtccagcgcgggtgcgtctcagctgcctgttaatccagcattttgagagctg 1740
Qy 1741 agcgggtggtacacctcaggtcaggtcaggttcaaaaaccagcctggcacaatagtgaaaccc 1800
Db 1741 agcgggtggtacacctcaggtcaggttcaaaaaccagcctggcacaatagtgaaaccc 1800
Qy 1801 cctctctactaaaaatgcacaaatattagtcggcgtggtgacactcctgttaatccacgc 1860
Db 1801 cctctctactaaaaatgcacaaatattagtcggcgtggtgacactcctgttaatccacgc 1860
Qy 1861 tacttaggagctgaggtgggaaatcaacttgactgcactcccaaggtgaggttcagtaagc 1920
Db 1861 tacttaggagctgaggtgggaaatcaacttgactgcactcccaaggtgaggttcagtaagc 1920
Qy 1921 tgaatcagctcagctcaccctagcttggtggtgcaagcaaaaactctataaaaaataa 1980
Db 1921 tgaatcagctcagctcaccctagcttggtggtgcaagcaaaaactctataaaaaataa 1980
Qy 1981 ttaataaattttcaaaaagctcgcgcaaaaataaaaaataaaaaataaaaaa 2040
Db 1981 ttaataaattttcaaaaagctcgcgcaaaaataaaaaataaaaaataaaaaa 2040
Qy 2041 aaaa 2044
Db 2041 aaaa 2044

```

RESULT 2

AAD02698

ID AAD02698 standard; cDNA; 2170 BP.

XX AC AAD02698;

XX DT 02-MAY-2001 (first entry)

XX DX Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.

Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT 5'UTR 9..343

FT FT /*tag= a

FT FT 344..1516

FT FT /*tag= b

FT FT /product= "Human glycosyl sulfotransferase-4alpha

FT FT (GST-4alpha)"

FT FT /note= "CDS is specifically claimed as SEQ ID NO: 4

FT FT in claim 6 (page no: 41) of the specification"

FT FT 1517..2134

FT FT 3'UTR

FT /*tag= c

XX WO200106015-A1.

XX PD 25-JAN-2001.

XX PF 19-JUL-2000; 2000WO-US19741.

XX PR 20-JUL-1999; 99US-0144694.

XX PR 13-JUL-2000; 2000US-0593828.

XX PA (RBSC) UNIV CALIFORNIA.

XX PI Rosen SD, Lee JK, Hemmerich S;

XX DR WPI; 2001-138471/14.

XX DR P-PSDB; AAY72639.

XX PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications -
Claim 6; Page 62; 128pp; English.

XX CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX CC GST is a type 2 membrane protein useful for inhibiting a binding event
between a selectin and a selectin ligand, which comprises contacting the
selectin with a non-sulphated selectin ligand, GST and a small molecular
agent that inhibits the sulphation activity of GST. GST is also useful
in inhibiting a selectin mediated binding event. GST is useful in gene
therapy to treat disorders such as acute or chronic inflammation, the
systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
dermatitis, myocarditis, regional enteritis, adult respiratory distress
syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
during transplantation.

XX SQ Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;

Query Match

Best Local Similarity 96.5%; Score 1973; DB 22; Length 2170;

Matches 1998; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

Qy 41 gaaagctccagtgaggaaactggtctcttgagagactctgtggtgcatagagtgattcaac 100
Db 167 gaaagctccagtgaggaaactggtctcttgagagactctgtggtgcatagagtgattcaac 226
Qy 101 caccctaaagaacacctggtctcttgagagactctgtggtgcatagagtgattcaac 160
Db 227 caccctaaagaacacctggtctcttgagagactctgtggtgcatagagtgattcaac 286
Qy 161 tgtgatcagctgttcagctctccagagacccctgagagctgtggtgcatagagtgattcaac 220
Db 287 tgtgatcagctgttcagctctccagagacccctgagagctgtggtgcatagagtgattcaac 346
Qy 221 tgggtgcaacggtttctccagcaagacagtgacagtgcctctctctggcacagacacctgc 280
Db 347 tgggtgcaacggtttctccagcaagacagtgacagtgcctctctctggcacagacacctgc 406
Qy 281 ctcctgtcttcctcatctcccgccagggccctccatcccccagccggtgagagatcgt 340
Db 407 ctcctgtcttcctcatctcccgccagggccctccatcccccagccggtgagagatcgt 466
Qy 341 gtccagctgtctgtctctctgtggtgagctgaggtcattctcttggccagctcttc 400
Db 467 gtccagctgtctgtctctctgtggtgagctgaggtcattctcttggccagctcttc 526
Qy 401 agccagcaccccgagctctctctacctgatggagccgctggtgcatgtgtgagacacccgt 460

Db	527	agccagcaccgcagcgtctctaccctgatggagccgcgtgcatgtggaaccacctg	586
Qy	461	tgcaggcgagcgcggcaacgctgcacatggcgtgcgcgacctgatgcgtctctatcttt	520
Db	587	tgcaggcgagcgcggcaacgctgcacatggcgtgcgcgacctgatgcgtctctatcttt	646
Qy	521	ttgtggacatgaacgtgtttgatgctctacatgccacagagcgaacctgtccgccttt	580
Db	647	ttgtgcacatggacgtgtttgatgctctacatgccacagagcgaacctgtccgccttt	706
Qy	581	ttcaactggcaacagaccgcgcgtgtgtctgcgccgcgcctgcagcgcctttccocga	640
Db	707	ttcaactggcaacagaccgcgcgtgtgtctgcgccgcgcctgcagcgcctttccocga	766
Qy	641	ggcaccatcaagcaagcaggcgtatgcaagacaactgtgcgcggcggcagccaattcaacctg	700
Db	767	ggcaccatcaagcaagcaggcgtatgcaagacaactgtgcgcggcggcagccaattcaacctg	826
Qy	701	gcccggagagcctgccgtcctacagcacagctggtgtctcaaggaggtgcgtcttctcaac	760
Db	827	gcccggagagcctgccgtcctacagcacagctggtgtctcaaggaggtgcgtcttctcaac	886
Qy	761	ctgcagggtgtctaccgctgtctcagcgaccccgcgtcaacctgcgcctcgtgcacctg	820
Db	887	ctgcagggtgtctaccgctgtctcagcgaccccgcgtcaacctgcgcctcgtgcacctg	946
Qy	821	gtgcgcgaccacggggccgtgtcgtgcgtcccgaggagcggcggcccgactctggcagcgc	880
Db	947	gtgcgcgaccacggggccgtgtcgtgcgtcccgaggagcggcggcccgactctggcagcgc	1006
Qy	881	gacaacgcactgtctgggcaccaacggcaagtgggtggaggcgcgacctcactgcgc	940
Db	1007	gacaacgcactgtctgggcaccaacggcaagtgggtggaggcgcgacctcactgcgc	1066
Qy	941	ctgattcgcgaggtgtgcgcgagccactgpgcattgcgcgagggcgcgcacactcaagccg	1000
Db	1067	ctgattcgcgaggtgtgcgcgagccactgpgcattgcgcgagggcgcgcacactcaagccg	1126
Qy	1001	ccacctctctgcgcgcgcgtctacgcgtgtgcgtcttgagacctggcgcgaggacgcg	1060
Db	1127	ccacctctctgcgcgcgcgtctacgcgtgtgcgtcttgagacctggcgcgaggacgcg	1186
Qy	1061	ctggcgagatccgcgcactctacgcttccacgcgcctgacccctcacgccacagctcgag	1120
Db	1187	ctggcgagatccgcgcactctacgcttccacgcgcctgacccctcacgccacagctcgag	1246
Qy	1121	gctggatccacaacatcacccacgggtcggggatcgccaagccaatcgaggccttccat	1180
Db	1247	gctggatccacaacatcacccacgggtcggggatcgccaagccaatcgaggccttccat	1306
Qy	1181	acttcgtctaggaatcgcgcaacgctctccaggcctggcgcaacgctgtgccttcaet	1240
Db	1307	acttcgtctaggaatcgcgcaacgctctccaggcctggcgcaacgctgtgccttcaet	1366
Qy	1241	aagatccctgcgctgacaggaggtgtgcgcgcgcgcctgcagctgtggttaccggcct	1300
Db	1367	aagatccctgcgctgacaggaggtgtgcgcgcgcgcctgcagctgtggttaccggcct	1426
Qy	1301	gtgtactctgcgacacagcagctgaacctcaacctggatctgtgtctgcacgagcccca	1360
Db	1427	gtgtactctgcgacacagcagctgaacctcaacctggatctgtgtctgcacgagcccca	1486
Qy	1361	gaccacttcagctggcgatcgctgactgagaaactctggccttagagcaagcccgaac	1420
Db	1487	gaccacttcagctggcgatcgctgactgagaaactctggccttagagcaagcccgaac	1546
Qy	1421	tgtgtgcgcagggcccaggaagcactgcattggtcgaaaaggagcctggcgcatggga	1480
Db	1547	tgtgtgcgcagggcccaggaagcactgcattggtcgaaaaggagcctggcgcatggga	1606
Qy	1481	a-caggtccactactacaacggaggttggttccctccctgaagtaagcaagagactgc	1539

Db	1607	agcaggctccctactatcaaacgggagtttgggtcctccctgaagttaggcaaggactgc	1666
Qy	1540	acgtttcttctctcctgattctcgttttctcttggattctctcgtgagctcctctcct	1599
Db	1667	acgtttcttctcctgattctcgttttctcttggattctctcgtgagctcctctcct	1726
Qy	1600	caggtgcaactctcatggaagcaactctgcacctcctctcttctggcgcaaggagta	1659
Db	1727	caggtgcaactctctcaggaagcaactctgcacctcctctctggtgcgaggagta	1785
Qy	1660	agttactgctaaattaaattaaattggtgcaggcgcggtgcggtggtcctgacctgtaa	1719
Db	1786	agttactgctaaattaaattaaattggtgcaggcgcggtgcggtggtcctgacctgtaa	1845
Qy	1720	tcccgacattttgagagctgaggcggtggatccactgaggtcaggattcaaaaccagc	1779
Db	1846	tcccgacattttgagagctgaggcggtggatccactgaggtcaggattcaaaaccagc	1905
Qy	1780	ctggcccaacatagtgaacccctctctactaaaaatgcataaaattagtcggcggtgtg	1839
Db	1906	ctggcccaacatagtgaacccctctctactaaaaatgcataaaattagtcggcggtgtg	1965
Qy	1840	gcacactcctgtaatcccgagctacttaggagctgaggtgggaatactctggactcca	1899
Db	1966	gcacactcctgtaatcccgagctacttaggagctgaggtgggaatactctggactcca	2025
Qy	1900	aaggtgaggtgtcagtagtgaatcattgcacatgcacctagcttgggtggcaaacg	1959
Db	2026	aaggtgaggtgtcagtagtgaatcattgcacatgcacctagcttgggtggcaaacg	2085
Qy	1960	aaactctatcaaaaaataattataaaattgttcataaagtcctgcgcaaaaaaaaa	2019
Db	2086	aaactctatcaaaaaataattataaaattgttcataaagtcctgcgcaaaaaaaaa	2145
Qy	2020	aaaaaataaaaaataaaaaa 2044	
Db	2146	aaaaaataaaaaataaaaaa 2170	
RESULT	3		
AAC76156			
ID	AAC76156	standard; cdna; 2988 BP.	
XX			
AC	AAC76156;		
XX			
DT	08-FEB-2001	(first entry)	
XX			
DE	Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.		
XX			
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;		
KW	vulnerable; antipsoriatic; antiparkinsonian; noctropic; neuroprotective		
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dermatological; antifungal; antineumatic; antiinflammatory;		
KW	antiviral; antibacterial; cancer; proliferative disorder; antithyroid;		
KW	antianaemic; gene therapy; osteoarthritis; graft vs host disease;		
KW	neurodegenerative disorder; diabetes mellitus; erythematous; infection;		
KW	cardiovascular disease; systemic lupus erythematosus; burn; wound;		
KW	cholesterol ester storage; immunodeficiency; malaria; autoimmune disorder; asthma		
KW	severe combined immunodeficiency; nocturnal haemoglobinuria; burn; wound;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;		
KW	thrombosis; contraceptive; ss.		
OS	Homo sapiens.		
XX			
PN	WO200058473-A2.		
XX			
PD	05-OCT-2000.		
XX			
PF	31-MAR-2000; 2000WO-US08621.		
XX			
PR	31-MAR-1999; 99US-0127607.		

XX PS Example 1; Page 62-104; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic

CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on

CC chromosome 16q23.1.

CC GST is a type 2 membrane protein useful for inhibiting a binding event

CC between a selectin and a selectin ligand, which comprises contacting the

CC selectin with a non-sulphated selectin ligand. GST and a small molecular

CC agent that inhibits the sulphation activity of GST. GST is also useful

CC in inhibiting a selectin mediated binding event. GST is useful in gene

CC therapy to treat disorders such as acute or chronic inflammation,

CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis

CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,

CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's

CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious

CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,

CC dermatitis, myocarditis, regional enteritis, adult respiratory distress

CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,

CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

XX during transplantation.

SQ Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

Query Match 85.8%; Score 1753.6; DB 22; Length 160552;

Best Local Similarity 99.1%; Pred. No. 2.5e-280;

Matches 1795; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

QY 197 cgggtgccccccagccgcgcagatggtggtccacaggttctccagcagacagtgacagtg 256

DB 47935 cgaagccccccagccgcgcagatggtggtccacaggttctccagcagacagtgacagtg 47994

QY 257 ctctctctggcagacacacacgtctctctgtctctctctctctctctctctctctctca 316

DB 47995 ctctctctggcagacacacacgtctctctgtctctctctctctctctctctctctctca 48054

QY 317 tccccagccgcccagagatggtgacgtggtggtggtggtggtggtggtggtggtggtggt 376

DB 48055 tccccagccgcccagagatggtgacgtggtggtggtggtggtggtggtggtggtggtggtg 48114

QY 377 tcatctcttctgggcccagctctctccagcagcaccccgagctctctctctctctctctct 436

DB 48115 tcatctcttctgggcccagctctccagcagcaccccgagctctctctctctctctctct 48174

QY 437 gcgtgcatggtgtgacacacacgtctgacggcagcgcgcgcgcgcgcgcgcgcgcgcgc 496

DB 48175 gcgtgcatggtgtgacacacacgtctgacggcagcgcgcgcgcgcgcgcgcgcgcgc 48234

QY 497 cgcacactgacgct 556

DB 48235 cgcacactgacgct 48294

QY 557 cagagccgaacactgtccgctcttctcaactggcgaacagcgcgcgcgcgcgcgcgcgcgc 616

DB 48295 cagagccgaacactgtccgctcttctcaactggcgaacagcgcgcgcgcgcgcgcgcgcgc 48354

QY 617 cccgctcgaagcgtcttcccgaggaacacatcagcaagcagcagcagcagcagcagcagc 676

DB 48355 cccgctcgaagcgtcttcccgaggaacacatcagcaagcagcagcagcagcagcagcagc 48414

QY 677 tgacgcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 736

DB 48415 tgacgcccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 48474

QY 737 ctcaaggaggtgcgtcttctcaactgagcagcagcagcagcagcagcagcagcagcagcag 796

DB 48475 ctcaaggaggtgcgtcttctcaactgagcagcagcagcagcagcagcagcagcagcagc 856

QY 797 ctcaactcgcacgtgcacgtggtgcgcgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 48534

DB 48535 ctcaactcgcacgtgcacgtggtgcgcgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 48594

QY 857 gggcgggcccgatctactggcagcgcagacacgcatcgtctggtggcaccacgcaagtgg 916

DB 48595 gggcgggcccgatctactggcagcgcagacacgcatcgtctggtggcaccacgcaagtgg 48654

QY 917 gtggaagccgacccctcaactgcgctgattcgcgaggtgtgcgcagcagcagcagcagc 976

DB 48655 gtggaagccgacccctcaactgcgctgattcgcgaggtgtgcgcagcagcagcagcagc 48714

QY 977 gccgagggcgcacactcaagccgacacccctctcgcggcgcgtacacgctggtggtg 1036

DB 48715 gccgagggcgcacactcaagccgacacccctctcgcggcgcgtacacgctggtggtg 48774

QY 1037 ttogaggaacctggcggggagccgctgcagacatccgcgcactctacgcttaccagcc 1096

DB 48775 ttogaggaacctggcggggagccgctgcagacatccgcgcactctacgcttaccagcc 48834

QY 1097 ctgacacctcaacgcacagcgtgagcgtggatccacacatccacccagcgtcggtggtg 1156

DB 48835 ctgacacctcaacgcacagcgtgagcgtggatccacacatccacccagcgtcggtggtg 48894

QY 1157 ggcaagccaatcgagcgtctccatctctctgtaggaatggcgccacacgtctccacgccc 1216

DB 48895 ggcaagccaatcgagcgtctccatctctctgtaggaatggcgccacacgtctccacgccc 48954

QY 1217 tggcgccacgcttgccttcaactaaatcctgcgctgcagaggtgtgcgcggcggtg 1276

DB 48955 tggcgccacgcttgccttcaactaaatcctgcgctgcagaggtgtgcgcggcggtg 49014

QY 1277 ctgacactgctgggctaccggcgtgtactctgcgacacagcagcgtgacctcaacctg 1336

DB 49015 ctgacactgctgggctaccggcgtgtactctgcgacacagcagcgtgacctcaacctg 49074

QY 1337 gatctggtgctgcacagcagccacacacacacacacacacacacacacacacacacac 1396

DB 49075 gatctggtgctgcacagcagccacacacacacacacacacacacacacacacacacac 49134

QY 1397 tgggcttagagcaagcccgacactgtgctgcagcagcccgaggaagcagcagcagcagc 1456

DB 49135 tgggcttagagcaagcccgacactgtgctgcagcagcccgaggaagcagcagcagcagc 49194

QY 1457 aaaaagagctggggcgccatggggaa-caggtccctactatcaacccggaggttgggtcc 1515

DB 49195 agaggagctggggcgccatggggaa-caggtccctactatcaacccggaggttgggtcc 49254

QY 1516 tccctgaagtaagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1575

DB 49255 tccctgaagtaagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 49314

QY 1576 agtcttctggagctgct 1635

DB 49315 agtcttctggagctgct 49373

QY 1636 tacctctctggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1695

DB 49374 tacctctctggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 49433

QY 1696 ggggtcggtggtcctatcct 1755

DB 49434 ggggtcggtggtcctatcctctctctctctctctctctctctctctctctctctct 49493

QY 1756 ctgaggtcagga-ttcaaaaaccagcctggccaacatagtgaaaacccctctctactaaa 1814

DB 49494 ctgaggtcagga-ttcaaaaaccagcctggccaacatagtgaaaacccctctctactaaa 49553

QY 1815 atgcaaaaattagtcggcggtggcgacactcctgtaatcccgactacttaggagcgtg 1874

DB 49554 atgcaaaaattagtcggcggtggcgacactcctgtaatcccgactacttaggagcgtg 49613

QY 1875 aggtgggaaattacactgacatcccaaggtgaggtgtgagtaagctgaatccac 1934

DB 49614 aggtgggaaattacactgacatcccaaggtgaggtgtgagtaagctgaatccac 49673

QY 1935 tgcacccctagctgggtggcagaaacactctatcaaaaaataataataatttgtt 1994

cc systemic lup

XX The present sequence is human glycosyl sulfotransferase-4beta (GSM-4
CC beta) cDNA. GSM-4 gene is found on chromosome 19q24.1.
CC GSM is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GSM and a small molecular
CC agent that inhibits the sulphation activity of GSM. GSM is also useful
CC in inhibiting a selectin mediated binding event. GSM is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis

[illegible]

```

Qy 1631 gccctactcttctggcgaggagtaagttaact 1666
    |||| | | | | ||| | | | |
Db 1525 gccacttcttcagagggcagccagctctact 1560

RESULT 7
AAZ20792
ID AAZ20792 standard; DNA; 2032 BP.
XX
AC AAZ20792;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human glycosyl sulfotransferase-3 coding sequence.
XX
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; human;
KW secondary lymph organ; ss.
XX
OS Homo sapiens.
XX
PN W09949018-AL.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
PA (REGC ) UNIV CALIFORNIA.
PA (SYNT ) SYNTAX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX
DR WPI; 1999-580442/49.
DR P-PSDB; AAY39918.
XX
PT Human and murine glycosyl sulfotransferase 3 and related
PS polynucleotides -
XX
PS Claim 4; Fig 1; 59pp; English.
XX
CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX
SQ Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;

Query Match 19.6%; Score 400; DB 20; Length 2032;
Best Local Similarity 63.6%; Pred. No. 1.3e-57;
Matches 645; Conservative 0; Mismatches 360; Indels 9; Gaps

QY 335 gatcgtgtgcacgtgtgtgtgtctctctgtggcgctggcctcctcttcttgggccag 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 gaggcatgcacgtgcgtgtgtctgtcttcttcttcttcttcttcttcttcttcttctt 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 395 ctcttcagcagcaccgcagcgtcttctacgtgatgagcccgctgtgcatgtgtggaacc 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 ctctttgggcagcaccagatggtttctacgtgatgagcccgctgtgcatgtgtggaacc 421
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 455 acctgtgcagggcagcgcggcagcgtgcacatggcgtgagcgaacctatgcgtct 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 accttcagcagcagcaccgcctggatgctgcacatggtgtgctggtatgctatacggtcc 481
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 515 atcttttctgacatgagcgtgtttgatgcctacatgccacag ---cggaaacctg 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Fri Feb 1 20:20:50 2002

(REGC) UNIV CALIFORNIA.
(SYNT) SYNTAX USA INC.

Bistrup A, Rosen SD, Tangemann K, Hemmerich S;

WPI: 1999-580442/49.

P-PSDB; AAT39919.

Human and murine glycosyl sulfotransferase 3 and related polynucleotides

claim 4; Fig 3; 59pp; English.

This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.

Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;

Query Match 17.3%; Score 354; DB 20; Length 1926;

Best Local Similarity 60.9%; Pred. No. 5e-50;

Matches 614; Conservative 0; Mismatches 385; Indels 9; Gaps 2;

594 -atcccaagaatgaatcatccccgggctcactcagctcctcagtcacagccc 552
691 attacgctggccgggggctcgcctctctacagcacgctggctcaaggagtgcg 750
653 tttagagttgttggaagcctgcctccctacagcagcgtgggtctcaaggagtgcg 712
751 cttcttcaacctgaggtgctacccgtgtcagcagcccgctcgaactgcgcat 810
713 cttcttcaacctgaggtgctacccgtgtcagcagcccgctcgaactgcgcat 772
811 cgtgacactgggtgagcagcccggtgctgctcctccgggagggcggtcccgat 870
773 cgtgacactgggtgagcagcccggtgctgctcctccgggagggcggtcccgat 832
871 actggcagcgacacagcagctgctgggcacacagcgcaagtgggtgagcgccacc 930
892 tctcatgattgacagtcgcatgttgatggggcagcagtcagcagaactcaagaaggaga 892
931 tcacctgcgctgattgcgaggtgtgcgcagccacagtcgcatcgccagggcgccac 990
893 ccaacctactatgtgatgaggtcatctgcgaagccagctggagatctacaagacat 952
991 actcaagccgcaccccttctgcggcgcctacccgctcagctgctcagggacctggc 1050
953 ccagctctgcccaagccctgcaggaacgtactctgttgctgctatgggacctggc 1012
1051 gcgggagccgtggaagatccgcgactctacgctctcccgccctgacccacgcc 1110
1013 tccagccctgtggccagactcccggaatgtatgaatcgtggatggaaattgtgcc 1072
1111 acagctcagcctgtgattcccaacatcacccaggggtgggagtcgccaagccaatcga 1170
1073 ccattctcagacactgggtgtataacatcacccagggcagggcgtga-----cca 1126
1171 ggccttcacatactgtctaggaatgcgcgaagctctccagcgcctgcgcacgcgtt 1230
1127 cgtttccacacaaatgccaggatgcccttaatgtctcccgaggtcgtgctgttt 1186
1231 gcccttcacatactgtctaggaatgcgcgaagctctccagcgcctgcgcacgcgtt 1290
1187 gcccttcacatactgtctaggaatgcgcgaagctctccagcgcctgcgcacgcgtt 1246
1291 ctaccgctgtgactctgctgggaccagcagcgtgacccactcactcactgactgtg 1345
1247 ctaccgctgtgactctgctgggaccagcagcgtgacccactcactcactgactgtg 1301

RESULT 9

AAZ20793

ID AAZ20793 standard; DNA; 1926 BP.

XX AC AAZ20793;

XX DT 08-DEC-1999 (first entry)

XX DE Mouse glycosyl sulfotransferase-3 coding sequence.

XX KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;

XX KW selectin binding interaction; inflammation; lymphocyte homing; mouse;

XX KW secondary lymph organ; ss.

XX OS Mus sp.

XX PN WO9449018-A1.

XX PD 30-SEP-1999.

XX PF 26-FEB-1999; 99WO-US04316.

XX PR 20-MAR-1998; 98US-0045284.

XX PR 12-NOV-1998; 98US-0190911.

XX

	Query Match	10.3%	Score 211;	DB 21;	Length 1043;
	Best Local Similarity	64.3%;	Pred. No. 1.7e-26;		
	Matches 205;	Conservative	53;	Mismatches 60;	Indels 1; Gaps 1;
QY	1671	aattaaattaaattgtgtgcagccgggtgcggtggtcctatgctgtgaatccacgattt	1730		
		: :			
Db	595	acaaacuuuuuuuuuugucagcgcgcggtggtgucacgucuguaucuccagcacuu	654		
		: :			
QY	1731	tgagagctgagcggtgtgatcacctgagtcagga-ticaaaacacgctgtgccaca	1789		
		:			
Db	655	ugggaggtgaggggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	714		
QY	1790	tatgtgaacccctctctactataaaatgcataatattgtccgggtgtgtgtgcaactct	1849		
		:			

[illegible]

RESULT	12
AAE21313	
ID	AAE21313 standard; DNA; 1601 BP.
XX	
XX	
XX	
AC	AAE21313;
XX	
DT	14-MAR-2001 (first entry)
XX	
XX	Human low adenosine antisense oligonucleotide related sequence #2880.
DE	
XX	
XX	Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW	human; airway disorder; bronchoconstriction; lung inflammation;
KW	surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW	immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW	respiratory obstruction; pulmonary obstruction; impaired respiration;
KW	asthma; asthmal. RDS;

KW	cancel, ss.
XX	
OS	Homo sapiens.
XX	
PN	WO2000062736-A2.
XX	
PD	25-OCT-2000.
XX	
PF	24-MAR-2000; 2000WO-US08020.
XX	
PR	06-APR-1999; 99US-0127958.
XX	
PA	(UYEC-) UNIV EAST CAROLINA.
PA	(NYCE/) NYCE J W.
XX	
PI	Nyce JW;
XX	
DR	WPI; 2000-679539/66.
XX	
PT	Low adenosine (A) content antisense oligonucleotides which do not
PT	trigger adenosine receptors during metabolism, useful e.g. for treating
PT	triglyceridemia - text obstructions -

XX Disclosure; Page 1316; 1592pp; English.
PS

PS Disclosure; Page 1316; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and/or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenously produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine molecules and their receptors, central nervous system (CNS) and peripheral nervous and non-nervous system receptors, CNS and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), pulmonary transplantation rejection, pulmonary infections, bronchitis, and/or cancer. AAR18434 to AAR21543 represent human polynucleotide fragments and antisense oligonucleotides used in the exemplification of the present invention.

Sequence 1601 BP; 445 A; 368 C; 356 G; 432 T; 0 other;

Query Match 10.3%; Score 211; DB 21; Length 1501;
 Best Local Similarity 80.9%; Pred. No. 1.7e-26;
 Matches 258; Conservative 0; Mismatches 60; Indels 1; Gaps 1.

[illegible]

RESULT 13
AAA35191
ID AAA35191 standard; DNA; 1601 BP.

XX
AC AAA35191;

XX
DT 28-JUL-2000 (first entry)

Human adenosine receptor r

Human adenosine receptor related polynucleotide 2nd SEQ ID NO:65.

KW	Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW	phosphorothioate; impaired respiration; inflammation; allergy;
KW	allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW	antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW	lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW	respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW	pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW	cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX	
XX	Homo sapiens.
XX	
XX	WO200009525-A2.
XX	
XX	24-FEB-2000.
XX	
XX	03-AUG-1999; 99WO-US17712.
XX	
XX	03-AUG-1998; 98US-0095212.
XX	
XX	{UYEC-} UNIV EAST CAROLINA.
XX	
XX	Nyce JW;
XX	
XX	WPI; 2000-205971/18.
XX	
XX	New antisense oligonucleotides useful for treating e.g. pulmonary
PT	vasoconstriction, inflammation, allergies, asthma, hypertension,
PT	bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT	cancers -
XX	
XX	
XX	Disclosure: Page 1232; 1343pp; English.
PS	
XX	

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airflow, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasize to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONs reduces side effects. The A-containing ONs break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2815, and then the last 195 sequences are also called SEQ ID NO:1 to 2815, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to AAA33992) are specifically claimed ONs from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing.

Sequence 1601 BP; 445 A; 368 C; 356 G; 432 T; 0 other;

Query Match 10.3%; Score 211; DB 21; Length 1601;
Best Local Similarity 80.9%;
Pred. No. 1.7e-26;
Matches 258; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

[illegible]

QY	1790	tagtgaaccccccctctctactaaataatgcgcaaatattagtcggcggttggtgcacacctcct	1849
Db	1273	tgttgaaccccgctctctactaaaaatcacaaaattgcggcgcatagttggcgcgcacct	1332
QY	1850	gtaatccccagctacttagaggctgagtggtgggaaaaatcaacttgactcccaaaagtgtgagg	1909
Db	1333	gtaatccccagctacttgaggagctggaggccaagagaattgcttgaccocccaggagatggaag	1392
QY	1910	ttgcagtagctgaataatcatgccactgcacctagcttgggtggcgaagcaaaactctat	1969
Db	1393	ttgcagtgagctgatcatcgtccctctgactccagcctgggtgacagagcaagactctgt	1452
QY	1970	caaaaaataattaataaa	1988
Db	1453	ctcaaaaaataaaaaataa	1471

RESULT	14
AAZ46814	standard; cDNA; 1601 BP.
ID	AAZ46814
XX	
AC	AAZ46814;
XX	
DP	31-MAR-2000 (first entry)
XX	
DE	Interleukin-10 (IL-10) cDNA sequence.
XX	
KW	Interleukin-10; IL-10; atopic dermatitis; allergic dermatitis;
KW	SLE; EB viral infection; lymphoma; antisense; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9967388-A1.
XX	
PD	29-DEC-1999.
XX	
PF	22-JUN-1999; 99WO-JP03315.
XX	
PR	24-JUN-1998; 98JP-0177188.
XX	
PA	(HISM) HISAMITSU PHARM CO LTD.
XX	
PI	Arima H, Tsuchiya S, Hirata T, Akiyama K, Goto T;
XX	
DR	WPI; 2000-106294/09.
XX	
PT	Antisense oligonucleotide for inhibiting expression of IL-10 protein,
PT	useful as active ingredient in remedies for atopic dermatitis -
XX	
PS	Disclosure; Page 19-20; 22pp; Japanese.
XX	
CC	The invention provides antisense oligonucleotides (AAZ46806-812) for
CC	inhibiting expression of interleukin-10 (IL-10) protein. A composition
CC	which is formulated with the antisense oligonucleotide or its
CC	derivatives as active ingredient can be used for treating atopic
CC	dermatitis, allergic dermatitis, SLE, EB viral infection or lymphoma. The
CC	present sequence represents a IL-10 cDNA sequence.
XX	
SQ	Sequence 1601 BP; 445 A; 368 C; 356 G; 432 T; 0 other;
XX	

```
Query Match      10.3%; Score 211; DB 21; Length 1601;
Best Local Similarity 80.9%; Pred. NO. 1.7e-26;
Matches 258; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
```

QY	1671	aattaaataaatgtgtgcgaagccgggtgctgtgctgaatcccaagcaatt	1730
DB	1153	acaaactaaattgtttcagccggcggtgctcagctgttaatcccaagcaatt	1212
QY	1731	tsgagagctgagcggtgtgatcacctgaggtcagga-ttcaaaacacagcctgtcccaaca	1789
DB	1213	tsgagagctgagcggtgtgatcaactgaggtcaggaaggttctcaaacacagcctgtgtaaca	1272

QY	1790	tatgtgaaaccccccctctctactataaaatgcacaaatagtccggcgttggtggcacactcct	1804
Db	1273	tggtagaaccccgctctctactaaaaatacaaaaattagccggcgtggtggcgcgaacct	1332
QY	1850	gtaatcccagctacttaggaggtcgtggtgggaaataccttgactccaaagtgtaggg	1909
Db	1333	gtaatcccgcgtacttggggagcgtgagccaagagaattgcttgaccaggagatggaag	1392
QY	1910	ttgcagtgaagctgaaatcatgcactgcacctgcaacctagcttgggtggcaagcaaaactctat	1969
Db	1393	ttgcagtgtagctgatactatgcctctgactccagcctgggtgcagagcaagactctgt	1452
QY	1970	caaaaaataataattataaa	1988
Db	1453	ctcaaaaaataaaaaataa	1471

RESULT 15

AAQ10207	standard; cDNA; 1618 BP.
ID	AAQ10207 standard; cDNA; 1618 BP.
XX	
XX	AAQ10207;
AC	
AC	19-MAR-1991 (first entry)
DT	
XX	pH5C insert containing Human cytokine synthesis inhibitory factor.
DE	
DE	Delayed-type hypersensitivity; DTH; leishmaniasis; parasite; MHC-associated autoimmune disease; interferon-gamma; ds.
XX	
XX	Homo sapiens.
OS	
XX	
XX	Key Location/Qualifiers
PH	33..566
FT	/*tag= a
FT	
XX	
XX	EP405980-A.
PN	
XX	
PD	02-JAN-1991.
XX	
XX	28-JUN-1990; 90EP-0307091.
PF	
XX	
XX	20-DEC-1989; 89US-0453951.
PR	
PR	28-JUN-1989; 89US-0372667.
XX	
XX	(SCHE) SCHERING CORP.
PA	
XX	
PI	Mosmann TR, Moore KW, Bond MW, Vieira PJM;
PI	
XX	
DR	WFI; 1991-009290/02.
DR	P-PSDB; AAR10158.
XX	
XX	Mammalian cytokine synthesis inhibitory factors - capable of inhibiting synthesis of cytokine(s) associated with delayed hypersensitivity and useful in treatment of e.g. leishmaniasis
PT	
PT	
PT	
XX	
XX	Disclosure; Fig 4; 31pp; English.
PS	
XX	
CC	The gene product may be used in treatment of diseases associated with MHC-linked immune response, suppressing a cell mediated humoral immune response. It may specifically be used to treat delayed-type hyper-sensitivity, leishmaniasis, and immune diseases
CC	
CC	
XX	
XX	Sequence 1618 BP; 461 A; 369 C; 356 G; 432 T; 0 other:
SQ	

```

Query Match      10.3%; Score 211; DB 12; Length 1618;
Best Local Similarity 80.9%; Pred. No. 1.7e-26;
Matches 239; Conservative 0; Mismatches 60; Indels 1; Gaps 1;
QY 1671 aattaaattaaatgtgcccagccgggtgcgggtgcctcatgctgaatcccagcattt 1730
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: January 31, 2002, 18:18:06
Job time: 9040 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:50:41 ; Search time 43.52 Seconds
 (without alignments)
 1310.804 Million cell updates/sec

Title: US-09-593-828-8
 Perfect score: 2051
 Sequence: 1 MWLPRFSSKVTVLLLAQTT.....LTDLVLRGPDHFSWASPD 390

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL17:*
 1: sp.archaea:*
 2: sp.bacteria:*
 3: sp.fungi:*
 4: sp.human:*
 5: sp.invertebrate:*
 6: sp.mammal:*
 7: sp.mhc:*
 8: sp.organelle:*
 9: sp.phase:*
 10: sp.plant:*
 11: sp.rodent:*
 12: sp.virus:*
 13: sp.unclassified:*
 14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2051	100.0	390	4 Q9UBY3	Q9UBY3 homo sapien
2	2051	100.0	411	4 Q9GZS9	Q9GZS9 homo sapien
3	1729.5	84.3	395	4 Q9GZX3	Q9GZX3 homo sapien
4	1536.5	74.9	395	11 Q9OUP4	Q9OUP4 mus musculus
5	1008	49.1	386	4 Q9V5R3	Q9V5R3 mus musculus
6	965.5	47.1	388	11 Q9RLI1	Q9RLI1 mus musculus
7	958.5	46.7	388	11 Q9WUE5	Q9WUE5 mus musculus
8	665	32.4	530	11 Q88276	Q88276 mus musculus
9	659.5	32.2	483	4 Q9GZNS	Q9GZNS homo sapien
10	659.5	32.2	530	4 Q9UEF2	Q9UEF2 homo sapien
11	656.5	32.0	484	4 Q9UED5	Q9UED5 homo sapien
12	656.5	32.0	531	4 Q9Y4C5	Q9Y4C5 homo sapien
13	617.5	30.1	484	11 Q9EPF78	Q9EPF78 mus musculus
14	614.5	30.0	484	11 Q99NB0	Q99NB0 mus musculus
15	610	29.7	486	4 Q75667	Q75667 homo sapien
16	609	29.7	486	4 Q9NS84	Q9NS84 homo sapien
17	577.5	28.2	479	4 Q75099	Q75099 homo sapien
18	576	28.1	472	11 Q88199	Q88199 mus musculus
19	560	27.3	474	11 Q9QZL2	Q9QZL2 rattus norv

20	520	25.4	411	11 Q9EQC0	Q9EQC0 mus musculus
21	513.5	25.0	411	4 Q43916	Q43916 homo sapien
22	472.5	23.0	441	13 Q93403	Q93403 torpedo cal
23	184.5	9.0	183	5 Q9VMC2	Q9VMC2 drosophila
24	174.5	8.5	363	5 Q9VMC4	Q9VMC4 drosophila
25	151	7.4	114	11 Q9DOK5	Q9DOK5 mus musculus
26	107.5	5.2	315	5 Q9VMK3	Q9VMK3 drosophila
27	99.5	4.9	972	5 Q9VLX0	Q9VLX0 drosophila
28	95	4.6	474	3 Q14192	Q14192 schizosacch
29	94	4.6	582	2 Q53639	Q53639 mycobacteri
30	94	4.6	933	5 Q9GNZ1	Q9GNZ1 leishmania
31	93.5	4.6	615	2 Q9RRS5	Q9RRS5 deinococcus
32	93	4.5	523	2 Q9I626	Q9I626 pseudomonas
33	92	4.5	925	1 Q9HLE8	Q9HLE8 thermoplas
34	91	4.4	920	2 Q9A0U7	Q9A0U7 streptococ
35	89.5	4.4	1844	12 Q56254	Q56254 turnip yell
36	88	4.3	417	2 Q83044	Q83044 azospirillu
37	87.5	4.3	661	5 Q9NKE9	Q9NKE9 drosophila
38	87.5	4.3	664	5 Q9VJW2	Q9VJW2 drosophila
39	87.5	4.3	937	2 Q9FB08	Q9FB08 streptomyc
40	87	4.2	620	2 Q9F3J2	Q9F3J2 streptomyc
41	87	4.2	632	5 Q9V3C9	Q9V3C9 drosophila
42	86.5	4.2	377	4 Q9H0V4	Q9H0V4 homo sapien
43	86.5	4.2	381	2 Q9WMM1	Q9WMM1 synchococc
44	86.5	4.2	445	10 Q9SZL2	Q9SZL2 arabidopsis
45	86.5	4.2	577	2 Q9RYK3	Q9RYK3 deinococcus

ALIGNMENTS

RESULT 1

Q9UBY3
 ID Q9UBY3 PRELIMINARY: PRT; 390 AA.
 AC Q9UBY3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.
 GN I-GLCNAC-6-ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=INTESTINE;
 RX MEDLINE=99423499; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
 Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL; AF176839; AAD56001.1;
 DR EMBL; AF176838; AAD56000.1;
 KW Transferase.
 SQ SEQUENCE 390 AA; 43832 MW; 132DIC3B6BE38179 CRC64;

Query Match 100.0%; Score 2051; DB 4; Length 390;
 Best Local Similarity 100.0%; Pred. No. 3.6e-175;
 Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPRFSSKVTVLLLAQTTCLLLFIISRPCSPSPAGGEDRVHVLVSSWSSGSSFLGQL 60

Db 1 MWLPRFSSKVTVLLLAQTTCLLLFIISRPCSPSPAGGEDRVHVLVSSWSSGSSFLGQL 60

Qy 61 FSPQPDVFLMEPAHWVMTLSQGSAAATLHMAVRDLMSIFLCDMDVFDAYMPSRNLSA 120

Db 61 FSPQPDVFLMEPAHWVMTLSQGSAAATLHMAVRDLMSIFLCDMDVFDAYMPSRNLSA 120

Qy 121 FFWNATSRALCSPACSAFFRGITISKQDVCKTICTROPFSLAREACRSYSHVVLKEVREF 180

Db 121 FFWNATSRALCSPACSAFFRGITISKQDVCKTICTROPFSLAREACRSYSHVVLKEVREF 180

```

181 NLQVLYPLSDPALNLRIVHLVRDPRVAVLSREAGPILARDNGIVLGTNGKWEADPHL 240
181 NLQVLYPLSDPALNLRIVHLVRDPRVAVLSREAGPILARDNGIVLGTNGKWEADPHL 240
181 NLQVLYPLSDPALNLRIVHLVRDPRVAVLSREAGPILARDNGIVLGTNGKWEADPHL 240
241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
301 EAMHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
301 EAMHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
301 EAMHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
361 PVSADQOQDRLDLVLPRGPDHFSWASPD 390
361 PVSADQOQDRLDLVLPRGPDHFSWASPD 390
361 PVSADQOQDRLDLVLPRGPDHFSWASPD 390

RESULT 2
Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (INTESTINAL N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).
DE ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).
DE CHST5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAG28023.1; -.
DR EMBL; AF219991; AAG26325.1; -.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match 100.0%; Score 2051; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.9e-175;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPFSSTKTVTLIAOQTCLLFIISRPSPAGGDRVHVLVLSWRSSTFLGQL 60
DB 22 MWLPFSSTKTVTLIAOQTCLLFIISRPSPAGGDRVHVLVLSWRSSTFLGQL 81
QY 61 FSOHPDVFYLMFAHVTWTLISQSAATLHMAVRDLMSIFLCLMDVDFAYMPCSRNLSA 120
DB 82 FSOHPDVFYLMFAHVTWTLISQSAATLHMAVRDLMSIFLCLMDVDFAYMPCSRNLSA 141
QY 121 FFWNATSRALCSPACSAFPRGTISKQDVKCTLCITRQPSLAREACRSYSHVVLKEVRF 180
DB 142 FFWNATSRALCSPACSAFPRGTISKQDVKCTLCITRQPSLAREACRSYSHVVLKEVRF 201

181 NLQVLYPLSDPALNLRIVHLVRDPRVAVLSREAGPILARDNGIVLGTNGKWEADPHL 240
181 NLQVLYPLSDPALNLRIVHLVRDPRVAVLSREAGPILARDNGIVLGTNGKWEADPHL 261
181 NLQVLYPLSDPALNLRIVHLVRDPRVAVLSREAGPILARDNGIVLGTNGKWEADPHL 261
241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
301 EAMHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
301 EAMHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
301 EAMHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLLGYR 360
361 PVSADQOQDRLDLVLPRGPDHFSWASPD 390
361 PVSADQOQDRLDLVLPRGPDHFSWASPD 411
361 PVSADQOQDRLDLVLPRGPDHFSWASPD 411

RESULT 3
Q9GZX3 PRELIMINARY; PRT; 395 AA.
AC Q9GZX3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE GST-4BETA (CORNEAL N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).
DE GST4BETA OR CHST6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal Localization and Genomic Organization for the Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-Sulfotransferase Gene Family.";
RL Glycobiology 0:0-0(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF280086; AAG48244.1; -.
DR EMBL; AF219990; AAG26325.1; -.
DR EMBL; AF219991; AAG26327.1; -.
KW Transferase.
SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 84.3%; Score 1729.5; DB 4; Length 395;
Best Local Similarity 85.8%; Pred. No. 2e-146;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
```

Db 180 NLOVLYPLSDPALNLRIVHLVRDPAVLRSGEQTAKALARDNGIVLGTNGTWVEADPGL 239
 QY 241 RLIREVCRSHVRIAEAAATKPPPLRGYRLVRVEDIAREPLAEIRALYATGTLTPQL 300
 Db 240 RVVEVCRSHVRIAEAAATKPPPLRGYRLVRVEDIAREPLAEIRALYATGTLTPQL 299
 QY 301 EAWIHNTHGSGIGKPTAEAFITSSRNARNVSOAWRHAFPTKILRVQVOCAGALQLLGYR 360
 Db 300 EAWIHNTHGSGIGKPTAEAFITSSRNARNVSOAWRHAFPTKILRVQVOCAGALQLLGYR 359
 QY 361 PVSADQORDTLDLVLPGRDPHFSWAS 388
 Db 360 PVSDEQRNALDLVLPGRNGFTWAS 387

RESULT 4
 Q9QUP4
 ID Q9QUP4 PRELIMINARY; PRT; 395 AA.
 AC Q9QUP4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.
 GN CHS75 OR I-GLCNAC-6-ST.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=INTESTINE;
 RX MEDLINE=99423499; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
 RT Sulfotransferase that is Highly Restricted to Intestinal Tissue."
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL; AF176841; AAD56003.1; -;
 DR EMBL; AF176840; AAD56002.1; -;
 DR MGD; MGI:1931825; Chst5.
 KW Transferase.
 SQ SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;

Query Match 74.9%; Score 1536.5; DB 11; Length 395;
 Best Local Similarity 76.0%; Pred. No. 3.6e-129;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MWLPFSSTVTVLLAQTCTLLFIIRSGPSSPAGGEDRVHVLVLSWSRSGSFLGOL 60
 Db 1 MKLPFSSTVMLSLMVQTG-ILVPLVRQVPSFAGLGERVHVLVLSWSRSGSFLGOL 59
 QY 61 FSQHPDVFYLMPEAWHVVTTLSQGSAAATLHMAVRDLMSIFLCMDVFDAYMPQSRNLSA 120
 Db 60 FSQHPDVFYLMPEAWHVVTTLSQGSAPALHMAVRDLMSIFLCMDVFDAYLPWRNTSD 119
 QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
 Db 120 LFQWAVSRALCSPVCEAFARGNTSSEVCKPLCATRPFGLAQEACRSYSHVVLKEVRF 179
 QY 181 NLOVLYPLSDPALNLRIVHLVRDPAVLRSGEQTAKALARDNGIVLGTNGTWVEADPGL 240
 Db 180 NLOVLYPLSDPALNLRIVHLVRDPAVLRSGEQTAKALARDNGIVLGTNGTWVEADPGL 239
 QY 241 RLIREVCRSHVRIAEAAATKPPPLRGYRLVRVEDIAREPLAEIRALYATGTLTPQL 300
 Db 240 RVVEVCRSHVRIAEAAATKPPPLRGYRLVRVEDIAREPLAEIRALYATGTLTPQL 299
 QY 301 EAWIHNTHGSGIGKPTAEAFITSSRNARNVSOAWRHAFPTKILRVQVOCAGALQLLGYR 360
 Db 300 QTWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQVOCAGALQLLGYR 359
 QY 361 PVSADQORDTLDLVLPGRDPHFSWAS 388
 Db 360 PVSDEQRNALDLVLPGRNGFTWAS 387

Db 360 SVHSELEQDLSLDLLPRGMDSEKWS 387

RESULT 5
 Q9Y5R3
 ID Q9Y5R3 PRELIMINARY; PRT; 386 AA.
 AC Q9Y5R3;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (L-SELECTIN LIGAND
 DE SULFOTRANSFERASE GST-3).
 GN GST3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TONSIL;
 RX MEDLINE=99264336; PubMed=10330415;
 RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
 RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
 RT "Sulfotransferases of two specificities function in the reconstitution
 RT of high endothelial cell ligands for L-selectin."
 RL J. Cell Biol. 145:899-910(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TONSIL;
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 RA Rosen S.D.;
 RT "Chromosomal Localization and Genomic Organization for the
 RT Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-
 RT Sulfotransferase Gene Family";
 RL Glycobiology 0:0-0(2001).
 DR EMBL; AF131235; AAD33015.1; -;
 DR EMBL; AF280088; AAG48246.1; -;
 DR InterPro; IPR003015; HLH_MYC.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 KW Transferase; Lectin; Selectin.
 SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 49.1%; Score 1008; DB 4; Length 386;
 Best Local Similarity 52.3%; Pred. No. 6e-82;
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 14 LLLAQTCTLLLTISR-
 Db 1 MLLPKMKLLFLVSQATLALFPFHMYSNLSLSMKAPFMRHVLVLSWSRSGSFLGQ 59
 QY 60 FSQHPDVFYLMPEAWHVVTTLSQGSAAATLHMAVRDLMSIFLCMDVFDAYM-POSRL 118
 Db 61 LFGQHPDVFYLMPEAWHVVTTKQSTAWLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 119 SFAFNWATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 178
 Db 121 SSLFQWNSRALCSPACDIIPQDEILIPAHRLCSPQPFVEVEKACRSYSHVVLKEVR 180
 QY 179 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 237
 Db 181 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 240
 QY 238 PHLRIVCRSHVRIAEAAATKPPPLRGYRLVRVEDIAREPLAEIRALYATGTLTPQL 296
 Db 241 QPYVYVQVICSQLEIYK--TIQSLKALQERYLVRYEDLARAPVAQTSRMYEVLGF 298
 QY 297 TPQLEAWIHNTHGSGIGKPTAEAFITSSRNARNVSOAWRHAFPTKILRVQVOCAGALQL 356
 Db 299 LPHQTWVHNITRGKMGD--HAFHTNARDALNVSQAWRHLPFAKIRRVQVOCAGALQL 356
 QY 357 LGYRPVVSADQORDTLDLVLPGRDPHFSWASPD 390
 Db 357 LGYRPVVSADQORDTLDLVLPGRDPHFSWASPD 390

Fri Feb 1 20:21:13 2002

us-09-593-828-8.rspt

Db 357 LGYRHVRSQEQNLLIDLL-----STWVPE 383

RESULT 6
Q9RI11 PRELIMINARY; PRT; 388 AA.

AC Q9RI11
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE L-SELECTIN LIGAND SULFOTRANSFERASE.

GN CHST4.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;

RN RN
RP SEQUENCE FROM N.A.
RX MEDLINE=99361934; PubMed=10435581;

RA Haraoka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis X, an L-selectin ligand displayed by CD34,"
RL Immunity 11:79-89(1999).

DR EMBL: AF109155; AAD45579.1; -;
DR MGD; MGI:1349479; Chst4.
KW Transferase; Lectin; Selectin.
SQ SEQUENCE 388 AA; 44635 MW; 6D5371AF6884AEE CRC64;

Query Match 47.1%; Score 965.5; DB 11; Length 388;
Best Local Similarity 55.9%; Pred. No. 3.8e-78;
Matches 194; Conservative 41; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVLLSSWRSGSFLGQLFQHPDVFYLMPEPAHVVHTLSQGSAAATLHMAVRLMSIF 101
DB 42 VHVLLSSWRSGSFLGQLFQHPDVFYLMPEPAHVVHTLSQGSAAATLHMAVRLMSIF 101

QY 102 LCDMDVFDAYM-PQSRNLSAFEKNWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFS 160
DB 102 LCDMSVFDAYMNPGRKQSSLFQWQSRALCSAPVCDFFPAHEISSPKHCKLCCGQPDF 161

QY 161 LAREACRSYSHVVLKEVRFENLQVLPILLSPALNLRIVHLVRDPAVRLSREAAGPILA 220
DB 162 MVKACRSHGVFLKEVRFSLQALYPLTDFSLNHLVHLVRDPAVRSRHTTIELV 221

QY 221 RDNIGVLGTNGKWEADPHRLIRE-----VCRSHVRIAE-ATLKPPPPFLRGY 269
DB 222 VDSHIVLG-----QHLETIKEEDQPYAMKIICKSQVDIVKAIQTL--PEALQORY 270

QY 270 RLVRFDLAREPLAEIRALYAFTGLTLTPOLEAWTHNTHGSGIGKPIEAFTSSRNARN 329
DB 271 LFLRYEDLVRLAPLAQTTRLYKFGVGLDFPLHLOTWVHNVTGKMGQ--HAFHINARNALN 328

QY 330 VSQAWRALPFTKILRYQVCEACAGALQLLGYRPVYSADQQRDLTLDLV 376
DB 329 VSQAWRWSLPYKYSQLODACEAMDLGLYQVRSQEQGNLSLDLL 375

RESULT 7
Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (CARBOHYDRATE (CHONDROITIN
DE 6/KERATAN) SULFOTRANSFERASE 4).
GN CHST4.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;

RN RN
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=TONSIL;
RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,
RA Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin,"
RL J. Cell Biol. 145:899-910(1999).

RN RN
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=TONGUE;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baidarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;

RA "Functional annotation of a full-length mouse cDNA collection,";
RL Nature 409:685-690(2001).

DR EMBL: AF131236; AAD33016.1; -;
DR EMBL: AK009113; BAB26078.1; -;
DR MGD; MGI:1349479; Chst4.

KW Transferase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 46.7%; Score 958.5; DB 11; Length 388;
Best Local Similarity 55.6%; Pred. No. 1.6e-77;
Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVLLSSWRSGSFLGQLFQHPDVFYLMPEPAHVVHTLSQGSAAATLHMAVRLMSIF 101
DB 42 VHVLLSSWRSGSFLGQLFQHPDVFYLMPEPAHVVHTLSQGSAAATLHMAVRLMSIF 101

QY 102 LCDMDVFDAYM-PQSRNLSAFEKNWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFS 160
DB 102 LCDMSVFDAYMNPGRKQSSLFQWQSRALCSAPVCDFFPAHEISSPKHCKLCCGQPDF 161

QY 161 LAREACRSYSHVVLKEVRFENLQVLPILLSPALNLRIVHLVRDPAVRLSREAAGPILA 220
DB 162 MVKACRSHGVFLKEVRFSLQALYPLTDFSLNHLVHLVRDPAVRSRHTTIELM 221

QY 221 RDNIGVLGTNGKWEADPHRLIRE-----VCRSHVRIAE-ATLKPPPPFLRGY 269
DB 222 VDSHIVLG-----QHLETIKEEDQPYAMKIICKSQVDIVKAIQTL--PEALQORY 270

QY 270 RLVRFDLAREPLAEIRALYAFTGLTLTPOLEAWTHNTHGSGIGKPIEAFTSSRNARN 329
DB 271 LFLRYEDLVRLAPLAQTTRLYKFGVGLDFPLHLOTWVHNVTGKMGQ--HAFHINARNALN 328

QY 330 VSQAWRALPFTKILRYQVCEACAGALQLLGYRPVYSADQQRDLTLDLV 376
DB 329 VSQAWRWSLPYKYSQLODACEAMDLGLYQVRSQEQGNLSLDLL 375

RESULT 8
Q88276 PRELIMINARY; PRT; 530 AA.
ID Q88276
AC Q88276;


```

DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
GN N-ACETYLGLOUCOSAMINE-6-O-SULFOTRANSFERASE LONG FORM.
OS ChST2.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-WHOLE EMBRYOS;
RX MEDLINE=98380482; Pubmed=9712885;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RL J. Biol. Chem. 273:22577-22583(1998).
DR EMBL; AB011452; BAA32138.1; -
DR EMBL; AB011452; BAA32139.1; -
DR EMBL; AB011451; BAA32137.1; -
DR MGD; MGI:1891160; Chst2.
KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;

Query Match 32.4%; Score 665; DB 11; Length 530;
Best Local Similarity 40.4%; Pred. No. 4.1e-51;
Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;

QY 37 GGEDRVHVLVLSWRSGLGFLGSHPDVYVIMPEPAHVVWTLTSGSAATHMAVRDL 96
Db 160 GGDKROLVYVFTTWRSGLGFLGSHPDVYVIMPEPAHVVWTLTSGSAATHMAVRDL 96
QY 97 MRSIFLDMVDVDFDAYMPQ---SRNLS--AFFNWTALRCLSPACSAFPGTISKOD--V 149
Db 220 LSALYRCDLSVQLYSPAGSGGRNLTGIFGAATNKVVCSSPLCPCAYRKEVGLVDRV 279
QY 150 CXTLCTROPFSLAREACRSYSHWLKEVRFNQLQVLYPSLPALNRLVHLVRDPRAVL 209
Db 280 CRK-CPPQRLARFECECRKYRTVWIKGVRFVAVLAPLKPDPALDKVHLVRDPRAVA 338
QY 210 RS-----REAAGPILARD-----NGVLGT---NGKWEADPH-LRLIREV 246
Db 339 SRSIRSGRLRESQVRSRDPRAHRMFFLEAGHKLGAKKEGGPADIHALGAMEVI 398
QY 247 CRSHVRIAEAPLKPFPFLRGYRLVRFDLAREPLAEIRALYAFGLTLTPQLEAWIHN 306
Db 399 CNSMAKTIQTA-LQPPDWLQGHLYVRVEDLVGDPVKTLRRYVDFGLVSPMEQFALN 457
QY 307 ITHGSG-IGKPTAEFTSSRNARNYSOAWRHALPFTKILRVOECAGALQLLYGRPVSA 365
Db 458 MTSGLSGSSSKP---FVVSARNATQAANAWRTALTFQOIKQVEEFCYQPMVILYGRVNSP 514
QY 366 DQORDLTLDLV 376
Db 515 EVKDLSTLL 525

RESULT 9
Q9GZN5 PRELIMINARY; PRT; 483 AA.
ID Q9GZN5
AC Q9GZN5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE N-ACETYLGLOUCOSAMINE 6-O-SULFOTRANSFERASE
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=PLACENTA;
RX MEDLINE=20498786; Pubmed=11042394;
RT Sakaguchi H., Kitagawa H., Sugahara K.;
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine 6-O-sulfotransferase that transfers sulfate to b-N-
RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence.";
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB021125; BAB16887.1; -
DR EMBL; AB021124; BAB16886.1; -
KW Transferase
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;

Query Match 32.2%; Score 659.5; DB 4; Length 483;
Best Local Similarity 39.18; Pred. No. 1.1e-50;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

QY 31 GPSSPAG-----GEDRVHVLVLSWRSGLGFLGSHPDVYVIMPEPAHVVWTL 81
Db 98 GVAAPPNGNTRGTGGVGDKROLVYVFTTWRSGLGFLGSHPDVYVIMPEPAHVVWTL 81
QY 82 SSGSAATHMAVRDLMSIFLDMVDVDFDAYMPQ---SRNLS--AFFNWTALRCLSPAC 136
Db 158 YPGDAVSLOGAARDMLSALYRCDLSVQLYSPAGSGGRNLTGIFGAATNKVVCSSPLC 217
QY 137 SAFPRGTISKQD--VCKTLCTROPFSLAREACRSYSHWLKEVRFNQLQVLYPSLPAL 194
Db 218 PAYRKEVGLVDRVCAK-CPPQRLARFECECRKYRTVWIKGVRFVAVLAPLDRPAL 276
QY 195 NLRVHLVRDPRAVRS-----REAAGPILARD-----NGVLGTNKWV 234
Db 277 DLKVIHLVRDPRAVRS-----REAAGPILARD-----NGVLGTNKWV 234
QY 235 --EADPH-LRLIREVCRSHVRIAEAPLKPFPFLRGYRLVRFDLAREPLAEIRALYAF 291
Db 337 GGPADYHALGAMEVICNSMAKTIQTA-LQPPDWLQGHLYVRVEDLVGDPVKTLRRYDF 395
QY 292 TGLTLTPQLEAWIHNTHGSG-IGKPTAEFTSSRNARNYSOAWRHALPFTKILRVOEVC 350
Db 396 VGLVSPMEQFALNMTSGSSSKP---FVVSARNATQAANAWRTALTFQOIKQVEEFC 452
QY 351 AGALQLLYGRPVYSADQORDLTLDLV 376
Db 453 YQPMVILYGRVNSPEEVKDLSTLL 478

RESULT 10
Q9Y6F2 PRELIMINARY; PRT; 530 AA.
ID Q9Y6F2
AC Q9Y6F2;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE CARBOHYDRAIE SULFOTRANSFERASE 2.
GN -CHST2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UMBILICAL VEIN ENDOTHELIUM;
RX MEDLINE=99168905; Pubmed=10049591;
RA Li X., Tedder T.F.;
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization.";
RL Genomics 55:345-347(1999).
DR EMBL; AF083066; AAD20981.1; -
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

```

Query Match 32.2%; Score 659.5; DB 4; Length 530;
Best Local Similarity 39.1%; Pred. No. 1.3e-50;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

QY 31 GPSSPAG-----GEDRVHVLSSWSGSGSFLGOLFSDHPDVFYLMPEAWHVVTTL 81
DB 145 GVAAPPNGRFGTGGVGRKQLVYFTWRSWGSFFGELFNQNPVEVFLYEPVHWQKL 204
QY 82 SQGSAATHMAVRDLMSIFLCMDVFDAYMPQ---SRNLS--AFFNWTASRALCSPAC 136
DB 205 YPGDAVSLOGAARDMLGALRYCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVCSPLC 264
QY 137 SAFPRGTISKOD--VCKTLCTROPFSLAREACRSYSHVVLKEVFFNLOVLYPLSDPAL 194
DB 265 PAYRKEVGLVDVDRCKK-CPPQRLARFEECRKYRTLVIRGVRVFDVAVLAPLRDPAL 323
QY 195 NLRVHLVDRPRAVRS-----REAGPILARD-----NGIVLGTNGKWV 234
DB 324 DLKVIHLVDRPRAVRSIRSRHGLIRSLQVRSRDRPRAHMPFLEAGHKLGAKEGV 383
QY 235 --EADPH-LRLIREVCRSHVRIAEATLKPPPLRGYRLVRFEDLAREPLAEIRALYAF 291
DB 384 GGPADYHALGAMVICNSMAKTIQTA-LQPPDWLQGHYLVRYEDLVGDPVKTLRRVYDF 442
QY 292 TGLMTLPQLEAWIHNTGSG-IGKPIEAFTSSRNARNVSOAHRHALPFTKILRVOEVC 350
DB 443 VGLIWSPEMEQFALNMTSGSSSKP---FVVSARNATQAANAWRTALTFOQIKOVEEFC 499
QY 351 AGALQLLGYRPVYSADQORDLTDLV 376
DB 500 YOPMAVLGYRVSPEEVKDLSTLL 525

RESULT 11
Q90ED5 PRELIMINARY; PRI; 484 AA.
AC Q90ED5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE (GLCNAC6ST).
GN GN6St.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsunaka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB014679; BAA34265.1; -
KW Transferase.
SQ SEQUENCE 484 AA; 53641 MW; D077EBCE7645F2F0 CRC64;

Query Match 32.0%; Score 656.5; DB 4; Length 484;
Best Local Similarity 39.7%; Pred. No. 2.1e-50;
Matches 149; Conservative 66; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGEDRVH-VLVLSWRSGSFLGOLFSDHPDVFYLMPEAWHVVTTLSGSAATLHMA 92
DB 110 APEGVGDKRHWMYFTTWRSWGSFFGELFNQNPVEVFLYEPVHWQKLPGDAVSLQA 169
QY 93 VRDLMSIFLCMDVFDAYMPQ---SRNLS--AFFNWTASRALCSPACSPRGTISKQ 147
DB 170 ARDMLSALRYCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVCSPLCPAYRKEVGLV 229

QY 148 D--VCKTLCTROPFSLAREACRSYSHVVLKEVFFNLOVLYPLSDPALNLRVHLVDRP 205
DB 230 DDRVCKK-CPPQRLARFEECRKYRTLVIRGVRVFDVAVLAPLRDPALDLKVIHLVDRP 288
QY 206 RAVLRS-----REAGPILARD-----NGIVLGTNGKWV--EADPH-LRL 242
DB 289 RAVASSIRSRHGLIRSLQVRSRDRPRAHMPFLEAGHKLGAKEGVGPGADYHALGA 348
QY 243 IREVCRSHVRIAEATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLPQLEA 302
DB 349 MEVICNSMAKTIQTA-LQPPDWLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 407
QY 303 WIHNITGSG-IGKPIEAFTSSRNARNVSOAHRHALPFTKILRVOECAGALQLLGYRP 361
DB 408 FALNMTSGSSSKP---FVVSARNATQAANAWRTALTFOQIKOVEEFCYQPNVILGYER 464
QY 362 VYSADQORDLTDLV 376
DB 465 VNSPEEVKDLSTLL 479

RESULT 12
Q9Y4C5 PRELIMINARY; PRI; 531 AA.
ID Q9Y4C5
AC Q9Y4C5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE LONG FORM OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE (GLCNAC6ST).
GN GN6St.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB014680; BAA34266.1; -
KW Transferase.
SQ SEQUENCE 531 AA; 58768 MW; D2EDH74E95B5162F CRC64;

Query Match 32.0%; Score 656.5; DB 4; Length 531;
Best Local Similarity 39.7%; Pred. No. 2.4e-50;
Matches 149; Conservative 66; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGEDRVH-VLVLSWRSGSFLGOLFSDHPDVFYLMPEAWHVVTTLSGSAATLHMA 92
DB 157 APEGVGDKRHWMYFTTWRSWGSFFGELFNQNPVEVFLYEPVHWQKLPGDAVSLQA 216
QY 93 VRDLMSIFLCMDVFDAYMPQ---SRNLS--AFFNWTASRALCSPACSPRGTISKQ 147
DB 217 ARDMLSALRYCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVCSPLCPAYRKEVGLV 276
QY 148 D--VCKTLCTROPFSLAREACRSYSHVVLKEVFFNLOVLYPLSDPALNLRVHLVDRP 205
DB 277 DDRVCKK-CPPQRLARFEECRKYRTLVIRGVRVFDVAVLAPLRDPALDLKVIHLVDRP 335
QY 206 RAVLRS-----REAGPILARD-----NGIVLGTNGKWV--EADPH-LRL 242
DB 336 RAVASSIRSRHGLIRSLQVRSRDRPRAHMPFLEAGHKLGAKEGVGPGADYHALGA 395
QY 243 IREVCRSHVRIAEATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLPQLEA 302
DB 396 MEVICNSMAKTIQTA-LQPPDWLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 454

QY	303	WIHNTHGSG-IGKPIEAFHTSSRNARVNSQARHALPFTKILRVOECAGALQILGYRP	361
Db	455	FALMTSGSGSSRP---FVVSARNATCAANARWALTFTQOIKQVEFCYQPMVLGYER	511
QY	362	VYSADQQRDLTDLV 376	
Db	512	VNSPEEVKDLTKLL 526	
RESULT	13		
Q9EPF78		PRELIMINARY;	
AC	Q9EPF78	PRT;	484 AA.
DT	01-MAR-2001	(TREMBLrel. 16, Created)	
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)	
DE	01-JUN-2001	(T-EMBLrel. 17, Last annotation update)	
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE		
DE	CHST7 OR GST5	(N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE)	
GN	CHST7		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J		
RX	PubMed=10936661;		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Yoshino M., Itoh M., Ishii Y.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka S.,		
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King H., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikatoo I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaigo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,		
RA	Hayashizaki Y.;		
RT	Functional annotation of a full-length mouse cDNA collection.;		
RT	Nature 409:685-690(2001).		
DR	EMBL; AF280089; AAC48247.1; -		
DR	EMBL; AB040710; BAB13769.1; -		
DR	EMBL; AK011202; BAB27465.1; -		
DR	MGD; MGI:1891767; Chst7		
DR	MGD; MGI:1919450; 2600013M07Rik.		
KW	transferease.		
QY	SEQUENCE	484 AA; 54766 MW; 9B195537D7AB7193 CRC64;	
QY	Query Match	30.1%; Score 617.5; DB 11; Length 484;	

Fri Feb 1 20:21:13 2002

QY 341 TKILRVOEVCAGALQLLGY 359
DB 441 EQVRQVEAACAPAMRLLAY 459

Search completed: January 31, 2002, 09:50:43
Job time: 331 sec

DB 263 RDEGLNLKVVQLPRDPRAVHNSRLKSRQGLRESIQVLRTRQGDHFHRYLLAHGVARP 322
QY 239 --HRLIREVCRSHVRIAEAA-----TLKPPPLFLRGYRLVRFEDLAREPLA 283
DB 323 GQARALPSAPRADFFLTSALEVICEAWLRLDLFTRCAPAWLRRYLRVYEDLVWQPOA 382
QY 284 EIRALYAFTGLITTPQLEAWIHNTHGSGIKGPIEAFTSSRNARNVSOAWRHALPFIKI 343
DB 383 QLRRLIRFSGLRTLAALDAFAFNTRGSAYGAD-RPFHLSARDAREAVHAWRERLSOEQV 441
QY 344 LRVOEVCAGALQLLGY 359
DB 442 RQVETACAPAMRLLAY 457

RESULT 15
O75667 PRELIMINARY; PRT: 486 AA.
AC O75667;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DJ7LL16.4 (N-ACETYLGLOUCOSAMINE-6-O-SULFOTRANSFERASE).
GN DJ7LL16.4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid-9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Uchimura K., Muramatsu T.;
RT "Identification and molecular characterization of a cDNA encoding a
novel N-acetylglucosamine-6-O-sulfotransferase";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL022165; CAA18154.1; -;
DR EMBL; AB040711; BAB13770.1; -;
KW Transferase.
SQ SEQUENCE 486 AA; 54266 MW; 3F1PDL430B3C8E95 CRC64;

Query Match 29.7%; Score 610; DB 4; Length 486;
Best Local Similarity 38.8%; Pred. No. 3e-46;
Matches 147; Conservative 52; Mismatches 126; Indels 54; Gaps 11;
QY 30 PGFSSPAGGE---DRVHVLVLSWRSGSFLGOLFQSPDPVFLMEPAWHVWTTLSQGS 85
DB 86 PSNLGAVGAEVSRKQHYVHATWRTGSSFLGELFNQHPDVFLYEPMMHLWQALYPGD 145
QY 86 AATLHMVAVRLMRSIFLCDDMDVFDAY-----MPQSRNL--SAFFNWTATRALCSPP 134
DB 146 AESLQALRDLRLSLRDFSVLRLYAPPDPAARAPDTAMITTAALFRFTNKVICSP 205
QY 135 ACSAFPRTGISQDVCKTCTIQ--PESL-AREA-CRSYSHVLKVEYRFFNLQVLYPLLS 190
DB 206 LCPGAPRAREVGLVEDTACERSCPPVAIRALEAECKYPVVVKDVRLLDLGLVLP 265
QY 191 DPALNLRIVHLVRDPA-----VLRREAAG-----PILARDNGIVLG 228
DB 266 DPGLNLKVVQLPRDPRAVHNSRLKSRQGLRESIQVLRTRQGDHFHRYLLAHGVARP 325
QY 229 TNGKWVADPH-----LRLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFEDLARE 280
DB 326 QCSRALPAAPRADEFLTGALEVICEAWLRLDLFARGA-----PAWLRRYLRVYEDLV 381
QY 281 PLAEIRALYAFTGLITTPQLEAWIHNTHGSGIGIKGPIEAFTSSRNARNVSOAWRHALP 340
DB 382 PRAQLRRLIRFSGLRTLAALDAFALNMTGAYGAD-RPFHLSARDAREAVHAWRERLSR 440

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: January 31, 2002, 09:46:11 ; Search time 25.96 Seconds
(without alignments)
1144.379 Million cell updates/sec
Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSKXIVTLLAQTT.....LTDLVLRGPDHFSWASPD 390
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	659.5	32.2	484	2 JE0261	N-acetylglucosamin
2	613.5	29.9	484	2 JC7350	N-acetylglucosamin
3	610	29.7	486	2 JC7351	N-acetylglucosamin
4	540.5	26.4	458	2 A57397	chondroitin 6-sulf
5	105	5.1	359	2 T16350	hypothetical prote
6	102.5	5.0	388	2 G70729	hypothetical prote
7	96	4.7	388	1 DEHPT	pyruvate dehydroge
8	96	4.7	707	2 S29029	phenylalanine ammo
9	95	4.6	474	2 T38905	probable valine--p
10	95	4.6	885	1 S78246	endopeptidase Clp
11	94.5	4.6	1581	1 VGNBJV	peplomer glycoprot
12	94	4.6	582	2 T0941	probable oxalyl-Co
13	93.5	4.6	615	2 C75278	DNA polymerase III
14	93	4.5	523	2 B83629	probable ATP-bindi
15	90.5	4.4	380	2 T42755	tyrosylprotein sul
16	86.5	4.2	445	2 T05639	hypothetical prote
17	86.5	4.2	577	2 B75585	probable long-chai
18	86.5	4.2	712	2 T33231	hypothetical prote
19	86	4.2	417	2 F71238	hypothetical prote
20	85.5	4.2	221	2 T50665	hypothetical prote
21	85.5	4.2	389	1 DEBQPA	pyruvate dehydroge
22	85.5	4.2	425	2 E83023	3-deoxy-D-manno-oc
23	85.5	4.2	571	2 S69210	protein kinase cak
24	85.5	4.2	1254	2 T41262	mutS family DNA mi
25	85	4.1	364	2 A49781	cholesterol dehydr
26	85	4.1	776	2 T31210	membrane protein t
27	85	4.1	1182	2 S14916	hypothetical prote
28	84.5	4.1	334	2 T44163	hypothetical prote
29	84.5	4.1	390	1 DERTP1	pyruvate dehydroge

30	84.5	4.1	390	1 DERTPA	pyruvate dehydroge
31	84.5	4.1	390	2 S23506	pyruvate dehydroge
32	84.5	4.1	491	2 A36036	cytochrome P450 2F
33	84.5	4.1	1844	1 ERWPTM	genome polyprotein
34	84	4.1	552	2 T44954	flagella-related p
35	84	4.1	590	2 T39246	protein phosphotas
36	83.5	4.1	390	1 DEHUPA	pyruvate dehydroge
37	83.5	4.1	420	2 A53531	concofetal trophobl
38	83.5	4.1	1844	2 S01956	hypothetical prote
39	83.5	4.1	1985	2 S19151	hypothetical prote
40	83.5	4.1	5149	2 F83345	probable non-ribos
41	83	4.0	370	2 A49360	pyruvate dehydroge
42	83	4.0	391	2 S23507	pyruvate dehydroge
43	83	4.0	590	2 T44416	protein phosphotas
44	83	4.0	1203	2 T04294	hypothetical prote
45	83	4.0	15281	2 S41309	cyclosporin synthe

ALIGNMENTS

RESULT 1
JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
R:Accession: JE0261
R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuo T.
J. Biochem. 124, 670-678, 1998
A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis
A:Reference number: JE0261; MUID:98391845
A:Accession: JE0261
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoad
C:Keywords: sulfotransferase

Query Match	32.2%	Score	659.5	DB	2	Length	484
Best Local Similarity	40.0%	Pred. No.	1.5e-51				
Matches	150	Conservative	67	Mismatches	121	Indels	37
Gaps	12						
QY	34	SPAGGEDRVH-VLVLSNRSGSSFLGOLFQHPDYFLMEPAHWVTTLSQGSAAATLHMA	92				
DB	110	APEGVGDRHWMYVFTWRSGSSFFGELFNQNPVEFFLYEPVHWVWQKLYPGDAVSLOGA	169				
QY	93	VRDLMRSTFLCDMDVFDAYMFQ---SRNLS--AFFNWTATSRALCSPACSAFFRGITISQK	147				
DB	170	ARDMLSALRYCDLSVFQLYSPAGSGGRNLTILGIFGATNKVVCSSPLCPAYRKEVGLV	229				
QY	148	D--VCKTILCTQPSLAREACRSHVVKVRFNLOVLYPLLSDPALNRIHLVRDP	205				
DB	230	DDRVCCK--CPPQRLARFECECRKYTLVKGVRVDVAVLAPLLRDPALDLKVHLVRDP	288				
QY	206	RAVLRS-----REAAGPILARD-----NCVIGTGKQW--EADPH-LRL	242				
DB	289	RAVASSRIRSHGLIRLSQVVRSDRPAHRMPFLEAAGHKLGAKKEGVGGPVDYHALGA	348				
QY	243	TREYCRSHVRIAEATLKPPPLFRGVRVLRVRFEDLAREPLAEIRALYAFGLTITPOLEA	302				
DB	349	MEVICNSMANTLIQTA-LQPPDWLQGHVLYVRYEDLVGEVPKTLRRVYDFVGLLVSPMEHQ	407				
QY	303	WIHNIHSGG-ICKPIBAFTSSNARNVSOAWRHALFFTKILRVQVVCAGALQLLGYRP	361				
DB	408	FALNMTSGSGSSSKP---FVVSARNATQAANWALTFTFOIKGVVEFCYQPMAYLOYER	464				
QY	362	VYSADOQDRLDLIV	376				
DB	465	VNSPEVKDKLSKTL	479				

```
RESULT 2
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7350
R:Uchimura, K.; Fasakhany, F.; Kadonatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
sulfate linkage formation; sulfate transport; sulfotransferase; transmembrane protein
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein
A:Reference number: JC7350; MUID:20374462
A:Accession: JC7350
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DDBJ:AB040710
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
sulfate linkage formation; sulfate transport; sulfotransferase; transmembrane protein
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein
Query Match 29.9%; Score 613.5; DB 2; Length 484;
Best Local Similarity 38.8%; Pred. No. 2.1e-47;
Matches 146; Conservative 54; Mismatches 127; Indels 49; Gaps 11;
QY 30 PGSPAG---GEDRVHVLVLSWRSGSFLGOLFSDHPDVFYLMPEAWHVVWTLISQGS 86
DB 85 PGNLSAAGEAVTQEKQHIYHATWRTGSSFLGELFNQHPDVFYLYEPMMHLMQALYPGDA 144
QY 87 ATLHMVRLMRSIFLCLDMVDVFDAY-----MPSRNLSEA--FFNWAISRALCSPPA 135
DB 145 ESLOGALRDLMLRSIFRCDFSVLRLYAOPGDPGERAPDSANLTAMLFWRWTKVICSPP 204
QY 136 CSAFPRGT----ISKQDVCKTLCTQFPESL-AREA-CRSYSHVVLKVEFFNQLVLYPLL 189
DB 205 CPAAPRADVGLVEDKACESTC--PPVSLRALEAECKRPVWIKDVLRLDGLVLPVLL 262
QY 190 SDPALNLRIVHVRDPRVAVLSREAGPILLARDNGIVLGTKCK-----WVEADP 238
DB 263 RDPGLNLKVVQLFRDPRAVHNSRLKSGRLRESIQVLRTRQGRDHFHRLVLAHGVADRP 322
QY 239 --HLRLREVCKSHVRIAEAA-----TLKPPFLRGYRLVRFEDLAREPIA 283
DB 323 GQARALSAFRADEFLLSALEVICEAWLRDLFTFGAPAWLRRLRYEDLVWQPOA 382
QY 284 EIRALYAFGLTLTPQLEAWIHNIHTHSGIGKPIEAFHTSSRNARNVSOAWHALPFTKI 343
DB 383 QLRLRLRTSGRLTAAALDAFAFNMTRGSAYGAD-RPFHLSARDAREAVHWRERLSRQEV 441
QY 344 LRQVEVCAGALQLLGY 359
DB 442 RQVETACAPAMRLLAY 457
RESULT 3
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R:Uchimura, K.; Fasakhany, F.; Kadonatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
sulfate linkage formation; sulfate transport; sulfotransferase; transmembrane protein
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein
A:Reference number: JC7351
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
sulfate linkage formation; sulfate transport; sulfotransferase; transmembrane protein
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein
Query Match 29.7%; Score 610; DB 2; Length 486;
```

```
Best Local Similarity 38.8%; Pred. No. 4.4e-47;
Matches 147; Conservative 52; Mismatches 126; Indels 54; Gaps 11;
QY 30 PGSPSPAGGE---DRVHVLVLSWRSGSFLGOLFSDHPDVFYLMPEAWHVVWTLISQGS 85
DB 86 PSNLISGAVGEAVSKEKHIVHAIWRTGSSFLGELFNQHPDVFYLYEPMMHLMQALYPGD 145
QY 86 AATHLMVAVRLMRSIFLCLDMVDVFDAY-----MQPSRNL--SAFENWATSRALCSP 134
DB 146 AESLOGALRDLMLRSIFRCDFSVLRLYAOPGDPAAARADPTANLTAAALFRWTKVICSPP 205
QY 135 ACSAFPRGTISKQDVCKTLCTQFPESL-AREA-CRSYSHVVLKVEFFNQLVLYPLL 190
DB 206 LCPGAPARAEEVGLVEDTACERSPPVAIRALEAECKRPVWIKDVLRLDGLVLPVLR 265
QY 191 DPALNLRIVHVRDPRVAVLSREAGPILLARDNGIVLGTKCK-----VLRSREAG-- 228
DB 266 DPGNLKVVQLFRDPRAVHNSRLKSGRLRESIQVLRTRQGRDHFHRLVLAHGVADRP 325
QY 229 TNGKWEADPH-----LRLIREVCKSHVRIAEAAATLKPPFLRGYRLVRFEDLARE 280
DB 326 GQARALSAFRADEFLLSALEVICEAWLRDLFTFGAPAWLRRLRYEDLVWQPOA 381
QY 281 EIRALYAFGLTLTPQLEAWIHNIHTHSGIGKPIEAFHTSSRNARNVSOAWHALPFTKI 340
DB 382 PRAQRLRLRTSGRLTAAALDAFAFNMTRGSAYGAD-RPFHLSARDAREAVHWRERLSR 440
QY 341 TKILRQVEVCAGALQLLGY 359
DB 441 RQVETACAPAMRLLAY 459
RESULT 4
A57397
chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (Chicken)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
C:Accession: A57397
R:Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kinata, K.; Shinomura, T.; Haba
J. Biol. Chem. 270, 18575-18580, 1995
A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotr
A:Reference number: A57397; MUID:95355490
A:Accession: A57397
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-458 <FUK>
A:Cross-references: GB:D49915; NID:9971262; PIDN:BAA08655.1; PID:9971263
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase
Query Match 26.4%; Score 540.5; DB 2; Length 458;
Best Local Similarity 34.3%; Pred. No. 7.6e-41;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;
QY 36 AGGEDRVHVLVLSWRSGSFLGOLFSDHPDVFYLMPEAWH--WTLISQGSAAATHMAY 93
DB 107 AAEPFRHVLMTATRTTGTSSFGVSEFFNQGNIFYLFEPLWHIERTVTFEGGANAVGSAL 166
QY 94 --BDLMRSIFLCLDMVDVFDAYM--POSRNLSAFENWATSRALCSPFACSPAPRGTSKQD 148
DB 167 VYKDLQQLLCDDLYILESTISPAPEBHLLTAALFRRGSSHSICEPEVCTPSLKKVFEKYH 226
QY 149 VCKTLCTQFPESLAREACRSYSHVVLKVEFFNQLVLYPLLSDPALMRLVHVRDPRV 208
DB 227 CKNERCGPLNTITLAAEACRRKQHMALKTVRTQLEFQPLAEDPRDLRIIQLVDRDPRV 286
QY 209 LRSEAAAPILARDNGIVLGTKWVE-----ADPHLRLIREVCKSHVRIAEAAATL 259
DB 287 LVSRMVA-----FSGKYESWKWAAEGEAPLQDEQVORLGNCEIS-IRLSAELGL 335
QY 260 KPPDFLRGRLVRFEDLAREPLAIFALYAFGLTLTPQLEAWIHNTTHSGIGKPIEA 319
```

Db	336	RQPEWLKRGTYMLVRYEDVAPARLKRKALEMYRFAGITHPTQVEWEIRANTQAP--QDSNG	392
QY	320	FHYSSNARNVSOAWRHAEFTKILRVQVSCAGALQLLGYRPVYSDAQQRDLTLDLVLPR	379
Db	393	IYSTQKNSSEQFEKWRFSIPFKLAQVQVQACFAMRLFGYKLASSAQETINSLSL-LEE	451
QY	380	GP 381	
Db	452	GP 453	

RESULT 5
T16350
hypothetical protein F42g9.8 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T16350
R:Taich, A.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of *C. elegans* cosmid F42g9.
A:Reference number: Z18498
A:Accession: T16350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-359 <TAI>
A:Cross-references: EMBL:U00051; NID:g1216305; PID:g1216308; PIDN:AAA91354.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F42G9.8
A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
C:Superfamily: *Caenorhabditis elegans* hypothetical protein F42G9.8

Query Match 5.1%; Score 105; DB 2; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.11;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

11	VTVLLAOTTCCLLFT-----ISRPGPSPPAGGEDRVH-----VLV 46
db	: : : : : : : : :
31	IYIFITCTCILLFISSIKCKKLEQLSLSKSFLTFNEQDAHSRLLNLNLEQLIFV 90
db	: : : : : : : : :
47	LSSWRSGSGFGLGOLFSDV-----FYMERPAW-HVWTTLSGCSAAATLHMVR 94
db	: : : : : : : :
91	GGVPESGTILRAILDADHPDVRCCGETMLLPSTFWQAGRNDW---NNSGIT-----141
db	: : : : : : : : :
95	DLMRSIFLCMDVDVFAYMPQSNNLSAFENWATSRLACSPPCSAFPRTGIS-KDVQCKTL 153
db	: : : : : : : :
142	-----QEVED-----DAVSATITEIVAKHSLSAPRL 167
db	: : : : : : : : :
154	CTROPFSLARACSYSHVIVKEVFFNWLYPLLSDPALNRIVHLVRDPRAVLRSR- 212
db	: : : : : : : : : : :
168	CNKDP-----YTALWLFTIR-----BLYP-----NAKFILMRDARAVVHSMI 205
db	: : : : : : : : : : : :
213	EAGOPILLARDNGIVLGTGKGWEADPHLRILEVCRSHVRTAEATLKPPFLRGRYLV 272
db	: : : : : : : : : : : :
206	ERKYPVAGYNISDLISMVQW---NOELRKMTFCQN---APGCIC-----V 247
db	: : : : : : : : : : : :
273	RFDLARREPIAEIRALYAFTGLTITPOLEAWIHNTHTSGTGKPI---EAFHTSS-RNA 327
db	: : : : : : : : : : :
248	YYERLIQKPABEILRIUNFLDLPESQQM-----LRHQDLLIGDEVOLDQEFASQVKNS 301
db	: : : : : : : : : : : :
328	RNVS--QAWRHALPFTKIILRVQEVCAQALQLGY-----RPVVS 364
db	: : : : : : : : : : : :
302	INTXALISWDPCFSEETLRKUDDV-AFPLGLTGYDTSLSKPDYS 344
db	: : : : : : : : : : : :

RESULT 6
:70729
:Hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37RV)
:Species: Mycobacterium tuberculosis
:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
:Accession: G70729
:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher C.; Harris, J.

#sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
#accession G07029
#author Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, R.; Rajadream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sqaes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A/Reference number: A70500; MUID:98295987
 A/Accession: G70729
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-388 <COL>
 A/Cross-references: GB:z77163; GB:AL123456; NID:g3261610; PTDN:CAB00968.1; PID:e25507
 A/Experimental source: strain H37Rv
 C/Genetics:
 A/Gene: Rv2267c

Query Match	5.0%; Score 102.5; DB 2;	Length 388;
Best Local Similarity	22.7%; Pred. No. 0.2;	
Matches	66; Conservative 37; Mismatches 99;	Indels 89; Gaps 15;
QY	44 VLVLSWRSGSSFLGOLF---SQH-----PDVYYLMEPAHWWTTLTSGSAAATH	90
Dd	: :: : :: : :	:
Dd	83 IFIVGHKRTGTTHLLHVVDVRTGTGYECLAPHFLLTE----WFAPYVEFLYSKH	137
QY	91 MAYVEDLMRSIFLCMDVFDAYCPQSRLNLSAFFENWATSRALCSPPACSAFRPGITSKDVC	150
Dd	:: : :: :	:
Dd	138 RAMDNMDLSL-----HHQEDE----FWC-MQGLESPLYLTIAFP-----	172
QY	151 KTLCTRQP-----FSIAREACRSYSHVLKEVRFFNLQVLY-----PLLSDDPALNLR-	198
Dd	173 ---NRPPQVEYLDELQAVPRELE--IKRKTLFRFQQYVFRKRKTIVILKNPTHFRFK	226
QY	199 -----VHLVRDPRAVLBSRAAGSTILARDNGIVL-----GTNGKWVEADPHLRLI	243
Dd	: :: :: :	:
Dd	227 VLLEVPPQAKFHIHRVPDYVVYPYSTIHHLKALRYHGLQOFTFDGLDKKVSVI--VVDLY	284
QY	244 REYCRSHRVIAEAATLKPPFELGRVLYRFEDFLAREPFAETCALYAFTGI,	294
Dd	285 R-----KLDEGREHVDPT-----REYERLEDLDGPGEGRRLYQHGLG	324

7
RESULT
DESCRIPT
pyruvate dehydrogenase (lipamide) (EC 1.2.4.1) alpha chain precursor, testis-specific
N:Alternate names: pyruvate dehydrogenase complex, E1 component alpha chain
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 15-Oct-1999
C:Accession: A37104
R:Daal, H.H.M.; Brown, R.M.; Hutchison, W.M.; Maragos, C.; Brown, G.K.
Genomics 8, 225-232, 1990
A:Title: A testis-specific form of the human pyruvate dehydrogenase E1alpha subunit 1
A:Reference number: A37104; MUID:91065637
A:Accession: A37104
A:Molecule type: mRNA
A:Residues: 1-398 <DA>
A:Cross-references: GB:M86808; GB:J04769; NID:g190789; PID:AAA60232.1; PID:g190790
Genetics:
A:Gene: GDB:PDHA2
A:Cross-references: GDB:120711; OMIM:179061
A:Map position: 4q22-q23
A:Superfamily: pyruvate dehydrogenase (lipamide) alpha chain; thiamin pyrophosphate-
Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; phosphoprote-
in; 1-27/Domain: transit peptide (mitochondrion) #status predicted <TNP>
28-368/Product: pyruvate dehydrogenase (lipamide) alpha chain #status predicted <M
183-232/Domain: thiamin pyrophosphate-binding domain homology <TPB>
230/binding site: phosphate (Ser) (covalent) #status predicted
291/binding site: phosphate (Ser) (covalent) #status predicted
298/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match	4.7%	Score 96;	DB 1;	Length 388;
Best Local Similarity	23.2%;	Pred. No. 0.76;		
Matches	45;	Conservative	25;	Mismatches 82; Indels 42; Gaps 7;

Fri Feb 1 20:21:11 2002

A: Experimental source: strain 972h-; cosmid c56E4

C: Genes: SPDB:SPAC56E4.03
A: Gene: SPDB:SPAC56E4.03
A: Map position: 1
C: Superfamily: Escherichia coli valine--pyruvate transaminase
C: Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate
F: 297/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 4.6%; Score 95; DB 2; Length 474;
Best Local Similarity 22.3%; Pred. No. 1.2;
Matches 69; Conservative 48; Mismatches 125; Indels 68; Gaps 17;

QY 33 SSPAGGEDRVHVLVLS-----WRSGSFLGQLFSQHDPDFVLMPEAHVW 78
DB 70 SSNSGKLDTVSPMSSSDVPLSVLQYGGSCALLSQFLKEHTRIIH--NPPFEGW 127
QY 79 T-TLSQGSAAHLHMAVRDL-----MRSIFL-----CDMDVFDAYMP 113
DB 128 NIITMTGNTSCDIALRMLTNRGDSILVEKYSFPLSQSMRPLGLSCIPIDMDQF-GFLP 186
QY 114 QSRNLSAFFNW-AFSRALCSPACSAFPGTISKQDVCKTLCTPQPSLAERACRSYHV 172
DB 187 ESMD-DILTNDWATSYGSPKPHVLYTTTGTGNTGTSILSVERRKQIYTLAK-----HDII 241
QY 173 VLKVEFFENLV-IV-----PLSDPALNLRIHVLRDPAVLRSREAGPILARON---GI 225
DB 242 ILEDPYVYLQMDAYEGKPEAAKFTNE--QFVKELIPSLSDMDVGRVIRMDLSLVK 299
QY 226 VLGTNGKWEADPHRLIREVCRSHVTAETATLKPPPLRG-RYKLYRF--EDLAREPL 282
DB 300 APGRVGVMTAOPLF-----IERGLRAETATQSASISQGLLYAMKHWGQDGYLEWL 353
QY 283 AEIRALYAF 292
DB 354 KHIR--YSYT 361

RESULT 8
S29029
phenylalanine ammonia-lyase (EC 4.3.1.5) - sweet potato
C: Species: Ipomoea batatas (sweet potato)
C: Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C: Accession: S29029
R: Tanaka, Y.; Matsuo, M.; Yamamoto, N.; Ohashi, Y.; Kano-Murakami, Y.; Ozeki, Y.
Plant Physiol. 90, 1403-1407, 1999
A: Title: Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase
A: Reference number: S29029
A: Accession: S29029
A: Molecule type: mRNA
A: Residues: 1-707 <TAN>
A: Cross-references: EMBL:M29232; NID:g168271; PIDN:AAA33389.1; PID:g168272
C: Superfamily: histidine ammonia-lyase
C: Keywords: ammonia-lyase; carbon-nitrogen lyase
F: 192-194/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F: 193/Modified site: dehydroalanine (Ser) #status predicted

Query Match 4.7%; Score 96; DB 2; Length 707;
Best Local Similarity 25.1%; Pred. No. 1.7;
Matches 57; Conservative 26; Mismatches 86; Indels 58; Gaps 10;

QY 159 FSLAREACRSYHVLEK-----VRFFMLQVLYPLLS---DPALNLRIVH 200
DB 131 FGNATESCHTLPHSATRAAMLVRINTILQYSGIRFEILEATKLLNHNITPCLPLR--G 188
QY 201 LVDPRAVLSREAGPILARDNGIVLGTNGKWEADPHRLIREVCRSHVRIAEAA--- 257
DB 189 TITASGDVPLSYTAGLITGRPSKAVGPNGETLNAEBALRL---SRSGRIRFVVASPR 244
QY 258 -----TLKPPPLRGYRLVRPDLAREPLAEI-RALYA-----FTGLITLTPLE--- 301
DB 245 KGLP-SLMAPPVLVGMASVLFANVLAVLSEVLSAIFAEVNGKPEFTD-HLTHKLKHP 303
QY 302 -----AMWNITHTSGSGIKGPTEAFH-----TSSRNARNVSOAW 334
DB 304 GQIEAAATMEHLIDGSSVKAQKLHEMDPLQPKQDRVALRTSPQW 350

RESULT 9
T38905
probable valine--pyruvate transaminase (EC 2.6.1.66) SPAC56E4.03 [similarity] - fission
C: Species: Schizosaccharomyces pombe
C: Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C: Accession: T38905
R: Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A: Reference number: Z21813
A: Accession: T38905
A: Status: translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-474 <CON>
A: Cross-references: EMBL:Z99261; NID:g4038621; PIDN:CAB16394.1; PID:g2414656; GSPDB:GN0D

Query Match 4.6%; Score 95; DB 1; Length 885;

[illegible]

RESULT 13

C75278
DNA polymerase III, tau/gamma subunit - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: C75278
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75278
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-615 <WH1>
A:CROSS-references: GB:AE002071; GB:AE000513; NID:G6460218; PIDN:AAF11953.1; PID:G646022
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2410
A:Map position: 1
C:Superfamily: DNA-directed DNA polymerase III gamma chain

Query Match 4.6%; Score 93.5; DB 2; Length 615;
Best Local Similarity 23.4%; Pred. No. 2.4;
Matches 92; Conservative 41; Mismatches 141; Indels 119; Gaps 21;
QY 4 PRESSKTVVLLAQTTCILLLFTISRPSPAGGEDRVHVLVLSWRSSSFLGQLFSQ 63
DB 44 PRGVGKTTIARLMTA-----NCTGPAPKPGCECECLAV-----RAGS----- 83
QY 64 HPDVFYLMFAHWHVITLSSGSAATLHMA--VRDLMSIFLCMD-----VEDAYMPQS 115
DB 84 HPDVMET-----DAASNNVDVDRDLREKVLGAAMRGKVIYILDEAHMS 129
QY 116 RNLSAFENWATSRALCPSPACAPPGCTISKQDVCKTLCRQPSLAREACRSY--SHVV 173
DB 130 R---AASN-ALLKLEPPHVFILATTEPEKIPIILSR-----CQHYFRRLT 176
QY 174 LKEY--RFFNLQVLYPLSDP-ALNL--RVH-LVRDPRVLRREAGPILARDNGIVL 227
DB 177 SEETAGLAGLVTEGASADPDALNLTGLADGAMRGESLLERMLAAGTAVTRP----- 231
QY 228 GTNCKWVEADPHRLIREVCVSHVRIEATLKPPTLGRYRVLRVEDLAREPLAETRA 287
DB 232 -----AVEEALGLPPEVRVGVASALLVGD-AGEAISGAQ 266
QY 288 LY--AFTGLTLPOLAWIHNITHGS-GIGKPIEAFHTSSRNARNVSQAWRHALPFTKIL 344
DB 267 LYRSGFARTVVEGLVAFGNALHAELGICE-----EGRLEGAEVPELL 310
QY 345 RVQEVACAGALLGYRPVYSADQORDYTLDLVL 377
DB 311 KIQ-----AALDEQAEFAARSADQO---SLELAL 336

RESULT 14
B83629
probable ATP-binding component of ABC transporter PA0136 [imported] - Pseudomonas aerugi
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83629
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: B83629
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <STO>

A:CROSS-references: GB:AE004451; GB:AE004091; NID:G9945958; PIDN:AAG03526.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0136

Query Match 4.5%; Score 93; DB 2; Length 523;
Best Local Similarity 26.4%; Pred. No. 2.1;
Matches 37; Conservative 17; Mismatches 60; Indels 26; Gaps 5;
QY 13 VLLAQTTCILLLFTISRPSPAGGEDRVHVLVLSWRSSSFLGQLFSQHPDVFYLM 72
DB 384 VLALAE-EIIRFAVKAPGAGAPARS-----LSGNLOKFLIGREILQAPLLVAAH 434
QY 73 PAHWVITLSSGSAATLH--MAVROLMRSTFLCDMDVFDAYMPQSRNLSAFFNWTASRA 129
DB 435 PTWCV---DVGAALHRLIALRDAGTALLVVSDELDELFLLSDR-----IAA 480
QY 130 LCSPPPACSAFPGTISKQDV 149
DB 481 LCSGRLLCPAVATAASAPQV 500

RESULT 15

T42755
tyrosylprotein sulfotransferase homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42755
R:Moore, K.L.; Ouyang, Y.B.
J. Biol. Chem. 273, 24770-24774, 1998
A:Title: Molecular cloning and expression of human and mouse tyrosylprotein sulfotran
A:Reference number: Z22262; MUID:98406128
A:Accession: T42755
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-380 <MOO>
A:CROSS-references: EMBL:AF049709; NID:G3617847; PIDN:AAC36062.1; PID:G3617848

Query Match 4.4%; Score 90.5; DB 2; Length 380;
Best Local Similarity 19.3%; Pred. No. 2.3;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;
QY 51 RSGSFLGQLFSQHPDV-----FYLMPEAW-----HWTTLSOG--SAATHNAV 93
DB 79 RSGTILRAMLDAHEVEVCGEETRVIPRIILNRSONKSEKWNRLQOAGVTGEVINNAI 138
QY 94 RDLMRSTFLCDMDVFDATMPQSRNLSAFFNWTASRALCPSPACSAFPGTISKQDVCKTL 153
DB 139 SSFIMEIM-----VHGDRAPRL 156
QY 154 CTROPFSLAREACRSYSHVLKEVRFNLQVLYPLSDPALNLRIVHLVRDPRVLRSE 213
DB 157 CNKDPFTMK-----SAVYLKE-----LFP-----NAKYLLMTRGQATVNS-- 192
QY 214 AAGPILARD--NGIVLG---TNGKWVEADPHRLIREVCVSHVRIEATLKPPTL 266
DB 193 ---IISRKVTITGFDLNDFOCMTKNAA---IQIMVDQCES---VGEKNCLK----- 236
QY 267 GRYLRVREDLAREPLAEIRALYAFGLTLPOLAWIHNITHGSGIGKPIEAFHTSSRN 326
DB 237 -----VYEQVLHPEACMRRTTELDI---FWDDKVLH---HEQLICKDISLNSVERSS 285
QY 327 ARNVSOA-----WRHALPFTKILRVQEVACAGALLGYRP 361
DB 286 DOVWKPVNLDALLKWKVGIPEVDVADMDSV-APMLRLGYDP 326

Search completed: January 31, 2002, 09:46:13
Job time: 156 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:51:05 ; Search time 16.97 Seconds
(without alignments)
842.621 Million cell updates/sec

Title: US-09-593-828-8

Perfect score: 2051

Sequence: 1 MWLPRFSKTVVLLAQIT.....LTDLVLPGRDHFWSWSPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	540.5	26.4	458	1	C6ST_CHICK
2	105	5.1	359	1	TPSB_CAEEL
3	102.5	5.0	388	1	YAG7_MYCTU
4	96	4.7	388	1	Q20351 caenorhabdi
5	96	4.7	707	1	P29803 homo sapien
6	95	4.6	885	1	P14166 ipomoea bat
7	94.5	4.6	1581	1	P49574 odontella s
8	90.5	4.4	380	1	P23052 berne virus
9	89	4.3	371	1	Q7081 caenorhabdi
10	88.5	4.3	249	1	Q9M174 laticauda c
11	86.5	4.2	376	1	P52994 rhizobium t
12	86.5	4.2	377	1	O68856 mus musculu
13	86.5	4.2	1608	1	O60704 homo sapien
14	85.5	4.2	221	1	P52639 borina disea
15	85.5	4.2	389	1	P46384 corynebacte
16	85	4.1	1182	1	P29804 sus scrofa
17	84.5	4.1	390	1	P15594 trypanosoma
18	84.5	4.1	390	1	P35486 mus musculu
19	84.5	4.1	491	1	P26284 rattus norv
20	84.5	4.1	1844	1	P24903 homo sapien
21	83.5	4.1	390	1	P08559 homo sapien
22	83.5	4.1	1844	1	P10358 turnip yell
23	83.5	4.1	1844	1	P28477 turnip yell
24	83	4.0	363	1	P52900 smintropsis
25	83	4.0	391	1	P35487 mus musculu
26	82	4.0	746	1	Q9KZV6 streptomyce
27	81.5	4.0	703	1	Q9EUS8 streptomyce
28	81.5	4.0	715	1	P11678 homo sapien
29	81.5	4.0	3011	1	P26654 h genome po
30	81	3.9	882	1	P10550 mycobacteri
31	81	3.9	972	1	Q29189 saccharomyc
32	80.5	3.9	704	1	O68045 rhodobacter
33	80.5	3.9	976	1	P21709 homo sapien

34	80.5	3.9	1942	1	Y054_HUMAN	P42694 homo sapien
35	80	3.9	247	1	NOH4_RHIME	P06237 rhizobium m
36	80	3.9	328	1	RPOA_NEINA	Q9J06 neisseria m
37	80	3.9	392	1	BM15_HUMAN	O95972 homo sapien
38	80	3.9	415	1	CXA3_RAT	P29414 rattus norv
39	80	3.9	416	1	CXA3_MOUSE	Q64448 mus musculu
40	80	3.9	933	1	ODOL_ECOLI	P07015 escherichia
41	79.5	3.9	559	1	MASY_MAIZE	P49081 zea mays (m
42	79	3.9	738	1	PLAK_XENLA	P30998 xenopus lae
43	78	3.8	462	1	TRPE_THETH	P05378 thermus aqu
44	77.5	3.8	370	1	TPS1_HUMAN	O60507 homo sapien
45	77.5	3.8	370	1	TPS1_MOUSE	O70281 mus musculu

ALIGNMENTS

RESULT 1	C6ST_CHICK	STANDARD;	PRT;	458 AA.
ID	C6ST_CHICK			
AC	Q92179;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17) (C6ST).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.			
RC	STRAIN-WHITE LEGHORN; TISSUE=Embryonic chondrocytes;			
RX	MEDLINE=93555490; PubMed=7629189;			
RA	Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,			
RA	Shinomura T., Habuchi O.;			
RT	"Molecular cloning and expression of chick chondrocyte chondroitin 6-			
RT	sulfotransferase.";			
RL	J. Biol. Chem. 270:18575-18580(1995).			
CC	!- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN			
CC	SULFATE.			
CC	!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLSULFATE + CHONDROITIN =			
CC	ADENOSINE 3',5'-BISPHOSPHATE + CHONDROITIN 6'-SULFATE.			
CC	!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE			
CC	(BY SIMILARITY).			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D49915; BAA08655.1; -			
KW	Transferrase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.			
FT	DOMAIN 1 23			
FT	TRANSMEM 24 37			
FT	DOMAIN 38 458			
FT	CARBOHYD 62 62			
FT	CARBOHYD 73 73			
FT	CARBOHYD 95 95			
FT	CARBOHYD 236 236			
FT	CARBOHYD 399 399			
FT	CARBOHYD 443 443			
SQ	SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;			

Query Match 26.4%; Score 540.5; DB 1; Length 458;
Best Local Similarity 34.3%; Pred. No. 3e-41;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;
QY 36 AGGEDRVHVLVLSWRSGSFTLGQIFSQHPDVFYLMPEPAHV--WTTLQSGSAAILHMAV 93

Query Match 5.1%; Score 105; DB 1; Length 359;
 Best Local Similarity 19.8%; Pred. No. 0.049;
 Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

107 RAPEPRHVLMMATRTGSGFEFFNOQGNIFLFEPLWHIERTVTFPPGGANAVGSAL 166
 QY 94 --RDLMSRSLFCDMDFDAYM--PQSRNLISAFENWATSRALCSPACSAFPRGTISKQD 148
 Db 167 VYRDVLQQLLCGLDLYLESFISAPAEHMTAALFRGSSHSLECEPVCPSLKKEVYKH 226
 QY 149 VCKTLCTROPFLAREACRSYSHVILKEVRFENLQVLYPLLSDPALNLRIVHVLVPRPV 208
 Db 227 CNKRCGPNLTLAAACRKHQWALKTVIRQLFELQPLAEDPRDLRLIQIQLVPRPV 286
 QY 209 LRSREAAGPILARDNGILVLTGNGKWE-----ADPHLRLEVECRSHVRIAEATL 259
 Db 287 LVSEMA-----FSKYESWKAAAGEAPLOEDEVORLGNCE--IRLSAELGL 335
 QY 260 KPPFLRGYELVREFDLAREPLAEIRALYAFGLITLTPOLEAWIHNITHGSGIGKPIEA 319
 Db 336 RQPRWLGRLVRYEDVVARPLKALEMYRFAGIHPTPOVEWIRANTQAP---QDSNG 392
 QY 320 FHTSSRNARNVSAWRHALPFTKILRVOECAGALOLLYGRPVVYVYVYVYVYVYVYVYV 379
 Db 393 IYSTOKSSSECFKWRSEIFPKLAQVYVODACEPAMRLFGYKCLASSAQELNRSLSL-LEE 451
 QY 380 GP 381
 Db 452 GP 453

RESULT 2
 ID TPSE CAPEL STANDARD; PRT; 359 AA.
 AC Q20351;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PUTATIVE PROTEIN-TYROSINE SULFOTRANSFERASE (EC 2.8.2.20)
 DE (TYROSYLPROTEIN SULFOTRANSFERASE) (TPST).
 GN F4269.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Taich A.;
 RL Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
 CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
 CC ACIDIC MOTIFS OF POLYPEPTIDES (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLISULFATE + PROTEIN TYROSINE =
 CC ADENOSINE 3',5'-BISPHOSPHATE + PROTEIN TYROSINE-O-SULFATE.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOFANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U00051; AAA91354.1; -;
 CC WormPep; F4269.8; CE07235.
 CC Hypothetical protein; Transferase; Transmembrane; Glycoprotein;
 CC KW Signal-anchor.
 CC DOMAIN 1 27 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC DOMAIN 49 359 LUMENAL, CATALYTIC (POTENTIAL).
 CC SEQUENCE 359 AA; 41469 MW; 9FC2F44539B8CC53 CRC64;

11 VIVLLAQITCLILFI-----TSRPGSPAGGEDRVH-----VLV 46
 Db 31 IYIFIEFCTLLIFSSIKCKLOEKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
 QY 47 LSSWRSSSLGOLFOSHPDV-----FYLMPAM-HVWTTLSQGSAAIEMAVR 94
 Db 91 GGPRSGTTLIRAILDAHPDVRCOGETMLPLSPLTWAGWENDWV-----NNSGIT----- 141
 QY 95 DLMRSIFLCMDVDPAYMQSRNLISAFENWATSRALCSPACSAFPRGTISK-ODVCKTL 153
 Db 142 -----QEVFD-----DAVSATFIFIAKHSLELAPRL 167
 QY 154 CTROPFLAREACRSYSHVILKEVRFENLQVLYPLLSDPALNLRIVHVLVPRPVRLR- 212
 Db 168 CNKDP-----YIALWLPTR-----RLYP-----NAKFLMRDARAVVHSMI 205
 QY 213 EAAGPILARDNGILVLTGNGKWEADPHLRLEVECRSHVRIAEATLKPFPPLRGYRIV 272
 Db 206 ERKVPVAGYNTSDELSMEVQW---NOELRKMTFOCNN---APGQCIK-----V 247
 QY 273 RFDLAREPLAEIRALYAFGLITLTPOLEAWIHNITHGSGIGKPI-----EAFPTSS-RNA 327
 Db 248 YVERLIQKPAEELIRITNFDLPFSQOM-----LRHQDLIGDEVLDNQDEFSASQVKN 301
 QY 328 RNVS--QAWRHALPFTKILRVOECAGALOLLY-----RPVYS 364
 Db 302 INTKALTSWDFCSFSEETLRKLDV-APFLGILGYDISISKPDYS 344

RESULT 3
 ID YM67_MYCTU STANDARD; PRT; 388 AA.
 AC Q50695;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 46.1 KDA PROTEIN RV2267C.
 DE RV2267C OR MT2329 OR MTCY339.43.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98235987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.

Fri Feb 1 20:21:12 2002

us-09-593-828-8.rsp

FIRST REACTION IN THE BIOSYNTHESIS FROM L-PHENYLALANINE OF A WIDE VARIETY OF NATURAL PRODUCTS BASED ON THE PHENYLPROPANE SKELETON.
 -!- CATALYTIC ACTIVITY: L-PHENYLALANINE = TRANS-CINNAMATE + NH(3).
 -!- PATHWAY: KEY ENZYME OF PHENYLPROPANOID METABOLISM.
 -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 -!- PTM: CONTAINS AN ACTIVE SITE 4-METHYLIDENE-IMIDAZOLE-5-ONE (MIO), WHICH IS FORMED AUTOCATALYTICALLY BY CYCLIZATION AND DEHYDRATION OF RESIDUES ALA-SER-GLY (BY SIMILARITY).
 -!- SIMILARITY: BELONGS TO THE PAL / HISTIDASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

 CC EMBL; M29232; AAA33389.1; .
 CC PIR; S29029; S29029.
 CC InterPro; IPR001106; PAL.
 CC Pfam; PF00221; PAL; 2.
 CC PROSITE; PS00488; PAL.HISTIDASE; 1.
 CC Lyase; Phenylpropanoid metabolism: Multigene family.
 CC SITE 192 194 MODIFIED TO FORM 4-METHYLIDENE-IMIDAZOLE-5-ONE (BY SIMILARITY).
 CC SEQUENCE 707 AA; 77138 MW; 87B9513A91228FBB CRC64;

Query Match 4.7%; Score 96; DB 1; Length 707;
 Best Local Similarity 25.1%; Pred. No. 0.76; Indels 58; Gaps 10;
 Matches 57; Conservative 26; Mismatches 86;
 QY 159 FSLAREACKSYSHVVKLE-----VRFNQLVLYPLLS---DPALNLRVH 200
 Db 131 FGNATESCHTLPSTRAMLVRLNTLOGYSIRFEILEATKLLNHNITPCPLR--G 188
 QY 201 LVDRPRAVLSREAAGPILARDNGIVLGTNGKWEADPHRLIREVCRSHVRAEAA--- 257
 Db 189 TITASGDIVPLSYIAGITGTRENSKAVGNGETLNAEALRL-----SRSGRIFRVASPR 244
 QY 258 ----TIKPPFRLGRYLRVFFEDLAREPLAEI-RALYA-----FTGLTLPQLE--- 301
 Db 245 KGLPSIMAPPLVGLMSVMFLFANVLAVLSEVLISALFAEVMGKPKFTD-HLTHKLKHP 303
 QY 302 ----AWHTNTTHSGIGKPIEAFF-----TSSRNARNVSOAW 334
 Db 304 GQIEAAIMEHILDGSSYVKAQKLEHMDPLQPKDQRYALRTSPQW 350

RESULT 6
 CLPC_ODOSI STANDARD; PRT; 885 AA.
 ID CLPC_ODOSI
 AC E49574; (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA HOMOLOG.
 GN CLPC
 OS Odontella sinensis.
 OG Chloroplast.
 OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
 OC Biddulphiophycidae; Eupodiscaceae; Odontella.
 OX NCBI_TaxID=2839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
 RT "The chloroplast genome of a chlorophyll a+c-containing alga,
 RT Odontella sinensis.";
 RL Plant Mol. Biol. Rep. 13:336-342(1995).
 CC -!- FUNCTION: MAY INTERACT WITH A CLPP-LIKE PROTEASE INVOLVED IN
 CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.

-!- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

 CC EMBL; Z67753; CAA91619.1; .
 CC InterPro; IPR003593; AAA.
 CC InterPro; IPR001939; AAA.subfam.
 CC InterPro; IPR001270; CLP_AB.
 CC Pfam; PF00004; AAA; 1.
 CC PRINTS; PR00300; CLPPROTEASEA.
 CC SMART; SM00382; AAA; 1.
 CC PROSITE; PS00870; CLPAB_1; 1.
 CC PROSITE; PS00871; CLPAB_2; 1.
 CC Chapterone; ATP-binding; Repeat; Chloroplast.
 CC NP_BIND 218 225 ATP (POTENTIAL).
 CC NP_BIND 560 567 ATP (POTENTIAL).
 CC SEQUENCE 885 AA; 99917 MW; 8F5BE76BSA65380B CRC64;

Query Match 4.6%; Score 95; DB 1; Length 885;
 Best Local Similarity 21.0%; Pred. No. 1.2; Indels 114; Gaps 18;
 Matches 89; Conservative 59; Mismatches 161;
 QY 5 RFSSKTVTVLLAQ-----TWCLLLFI--SRGPSPAGGEDRVHVLVSSWR 51
 Db 4 KFTGAIKVIMLSQDEARRMGHNFVGTQQLLIGIQRHGIGARALKKQV---TLKKAR 60
 QY 52 -----SGSPLGQLFSQHPDVFYLMEPANHVWTLTSGSAAATHMAVRLMRS--IF 101
 Db 61 REIELYIGRGTFVASEIPTPRAKRVLEMAVHEGKDLGONFYGTETHEILLALISESDGVA 120
 QY 102 LCDMDVFDAYMPOSRLSAFFNWTASRALCSPACSAFFPGTISQDVCKTCTROPFSL 161
 Db 121 MRULDKLGVNPKRLNLIYI-----ENQEEILRLPLQAEKFI 161
 QY 162 ARBACHS-----YSHVVLKVEFFNQLVLYPLLSPPALNL-RIVHLVRDRAVLRSRE 213
 Db 162 EREKKGSSPTLDEYSENISKEAVDGKL-----DPVIGRDKIEHV--IKVLARRR 210
 QY 214 AAGPILARDNGIVLGTNGKWEADPHRLIREVCRSHVRAEATLKPDPFLGRYLRV 273
 Db 211 KNPVLIGEPGV-----GKTAVAEGLAQLI-----IAE---KAPDFLDGNL--- 248
 QY 274 FEDLAREPLAEIRALYAFTGLTLPQLEAWHTNTHSGIGKPIEAFF-----SSRNAR 328
 Db 249 ---LMALDLSILLAGTKYRG-EFEERIKRIVEEVONDSAILLYDEHTLVGAGAAGAV 304
 QY 329 NVSQAWRHALPFTKILRVQFVCAGALQLGYR-----PVYSADQQRDLTLD 374
 Db 305 DAANILKPALARGKFR-----CIGATTIDEYKRIEDPALERFQPVHVKEPFGVTIE 359

RESULT 7
 VGLP_BEV STANDARD; PRT; 1581 AA.
 ID VGLP_BEV
 AC P23052;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-NOV-1991 (Rel. 20, Last annotation update)
 DE PEPLIMER GLYCOPROTEIN PRECURSOR.
 GN P.
 OS Berne virus (BEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

Coronaviridae; Torovirus.
 NCBL_taxid=11156;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=ISOLATE PL38/72;
 RX MEDLINE=91020973; PubMed=2219698;
 RA Snijder E.J., den Boon J.A., Spaan W.J.M., Weiss M., Horzinek M.C.;
 RT "Primary structure and post-translational processing of the Berne
 virus peplomer protein.";
 RL Virology 178:355-363(1990).
 CC -!- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS
 CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52506; CAA36748.1; -
 DR PIR; A36759; VGWJBV.
 KW Glycoprotein; Envelope protein; Transmembrane; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 1581 PEPLIMER GLYCOPROTEIN.
 FT TRANSMEM 1547 1572 POTENTIAL.
 FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1267 1267 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1297 1297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1385 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1389 1389 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1428 1428 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1431 1431 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1438 1438 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1483 1483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1495 1495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1515 1515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1591 1591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1591 AA; 178332 MW; 00D91B41837AC769 CRC64;
 Query Match 4.6%; Score 94.5; DB 1; Length 1581;
 Best Local Similarity 17.6%; Pred. No. 2.9; Mismatches 135; Indels 179; Gaps 19;
 Matches 81; Conservative 65;
 2 WLPRESSKTVIVLLAQTCLLLFTISRPSPAGGEDRVHVLVLSWRSGSFLGOLF 61
 259 WVAFQNKATAVILFSELVFAQVTR- - - - -RLG 289
 62 SQHPDVFVLMPPAHVWTLTSGSNAILHMAVRLMRSIFLCMDVFDAYMPQSRNLSAF 121
 290 VNTDFVFWLVKQAH- - - - -YLSQA-NLSP- 313
 122 FNAWTSRALCPAPACAPRGTGSKQDVCKTLCTRPESLAREACRSYSHVV-LKEVRF 180
 314 -NYALFSAICN- - - - -SLYQGSATLSTLCFSPFFVAQEC- - -YNNALYLPDAVFT 360
 181 NL- - - - -QVLYPL- - - - -LSDPALNI- - - - -RIVHLVRDPRVAVL- - - - - 209
 361 TLFSTLFSWDYQINPLNVLTONETFLQPATNYQGTLSQGRMLNFKDAIVLDFD 420
 210 - - - - -RSREACP- - - - -ILARDNGVLGTNGKWEADPELRILREVCRSHVRIASATLKP 261
 421 TKFRTNDAPSSDIFVFWVAQAQLIRYGNFRIEQINGYFQV- - - - -KCSSNI- - - - -ISTLEP 473
 262 PFLRGYRLVRPDLAREPLAEATRYLYAFT- 292

Db 474 HP--AGVIMIAHRSWMSVAARNSTSYFCVTHSLTTEGKLDISISWFFHTLALPSGPVSQ 531
 QY 293 -----GLTLTPOLEAWTHNLT-----HSGGCKGPIEAPHTTSRRNARNS 331
 Db 532 VSMPELLSTAAGVGYMHPMTEHPIPLTLTQAOSQYQPSFFNIGINKTI-TLTTLQAYAAQVY 590
 QY 332 QAWRUALPTFKILRVQEVOCAGALQLLGYRPVYSADQORDL 371
 Db 591 TAWFLSVIVVRLPEARRLTLG-VQLVPFTQALLSIKQADL 629
 RESULT 8
 TPISA_CAEEL
 ID TPISA_CAEEL STANDARD; PRT; 380 AA.
 AC 077081; Q9NEW9;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE SULFOTRANSFERASE A (EC 2.8.2.20) (TYROSYLPROTEIN
 DE SULFOTRANSFERASE-A) (TPST-A).
 GN Y111B2A.15.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peleoderinae; Caenorhabditis.
 OX NCBL_taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=98406128; PubMed=9733778;
 RA Ouyang Y.-B., Moore K.L.;
 RT "Molecular cloning and expression of human and mouse tyrosylprotein
 RT sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in
 RT Caenorhabditis elegans.";
 RL J. Biol. Chem. 273:24770-24774(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Sulston J.E.;
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Durbin R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
 CC ACIDIC MOTIFS OF POLYPEPTIDES.
 CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + PROTEIN TYROSINE =
 CC ADENOSINE 3',5'-BISPHOSPHATE + PROTEIN TYROSINE-O-SULFATE.
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF049709; AAC36062.1; -
 CC EMBL; ALI32904; CAC35844.1; -
 CC WormPep; Y111B2A.15; CE26632.
 KW Transferrase; Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT FT (POTENTIAL).
 FT DOMAIN 28 380 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 380 AA; 43313 MW; FF709BF00FEDC95 CRC64;
 Query Match 4.4%; Score 90.5; DB 1; Length 380;
 Best Local Similarity 19.3%; Pred. No. 1.1;

[illegible]


```

CC EMBL; U4722; AAB08982.1; -
DR EMBL; X87608; CAA60912.1; -
KW Modulation; Transferase.
SQ SEQUENCE 249 AA; 28510 MW; C881E1045B9F5940 CRC64;

Query Match
Best Local Similarity 4.3%; Score 86.5; DB 1; Length 249;
Matches 65; Conservative 54; Mismatches 108; Indels 117; Gaps 16;

QY 26 IISRPGPSAGGEDRVHVLVLSNRSSGSLGQLFSCHPDVFYLMPEAHVWVTLSQGS 85
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 2 IHSLSPP-----FVILAPRTGTHYLEALNDHPNI-----LSNG- 38

QY 86 AATLHMAVRDLMSIFLCLMDVFDAYMQSNLSAFENWATSRALCSPACSPACSPAC 145
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 39 -----ELLNS-----YDENPDKDLRH-----S 57

QY 146 KODYCKTLCITROPFLAREACRSYSHVVK--EVRFFNLQVLYPLSDPALNLRIVHLVR 203
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 58 DRELELAYMRYP-----PAKKVTHGVCKINEPQYERPGFFDELA-RWQGLKVLILIR 111

QY 204 DPAVLRSREAAGPILARDNGIVLTGSKVE-----ADPHLRILREVCRS----- 249
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 112 NTLESLSL-----VQARQ-----TGQWLKFPSPDRNEPPSVLSVNECEAYEKAADDF 159

QY 250 HVRIAEAAATLPPFLGRYRIVRFEDLAREPLAEIRALYAFTGLTILTPQLEAMHNIH 309
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 160 HARVKDA-----FDPKLMTEYQDLLKPSACIAAVLAFLG-----APAHRFNS 204

QY 310 GSGIGKPIEAFTSSRNARVNSQAHAL--PDTKILRVOEVCA 351
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 205 RAVIQR--QETRLARSLRNFVELRRHFAGGPGYAKFFELDDASA 246

RESULT 11
TPS2_MOUSE
ID TPS2_MOUSE STANDARD; PRT; 376 AA.
AC O8856;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROTEIN-TYROSINE SULFOTRANSFERASE 2 (EC 2.8.2.20) (TYROSYLPROTEIN
SULFOTRANSFERASE-2) (TPST-2).
GN TPST2 OR DSUCLA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98406128; PubMed=9733778;
RA Ouyang Y.-B., Moore K.L.;
RT "Molecular cloning and expression of human and mouse tyrosylprotein
sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in
Caenorhabditis elegans.";
RL J. Biol. Chem. 273:24770-24774(1998).
CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
ACIDIC MOTIFS OF POLYPEPTIDES.
CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + PROTEIN TYROSINE =
ADENOSINE 3',5'-BISPHOSPHATE + PROTEIN TYROSINE-O-SULFATE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

CC EMBL; AF049890; AAC36060.1; -
DR MGD; MGI:1309516; Tpst2.
KW Transferase; Transmembrane; Glycoprotein; Signal-anchor;
KW Multigene family.
FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 9 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 26 376 LUMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 376 AA; 42067 MW; E714C7DE3D8718D0 CRC64;

Query Match
Best Local Similarity 4.2%; Score 86.5; DB 1; Length 376;
Matches 73; Conservative 36; Mismatches 112; Indels 119; Gaps 18;

QY 51 RSGSFIQLFSGHPDV-----FYLMEPAHVWVTLSQGSAAATLHMA-VRDLMR 98
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 77 RSGTFLMEALDAHDEVRCGETRIIPRVLMRQAW---TKSGREKRLDEAGVTD--- 129

QY 99 SIFLCLMDVFDAYMQSNLSAFENWATSRALCSPACSPACSPACSPACSPACSPAC 157
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 130 -----EVLDA-----AMQAFILEVIAKHGEPARVLCNKD 158

QY 158 PFSLAREACRSYSHVILKEVRFENLQVLYPLSDPALNLRIVHLRVPRAVLRSRAAGP 217
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 159 PFTLK-----SSVYLAR-----LFP-----NSKFLLMVRGDSVSHVSNITRKV 196

QY 218 ILARDNGIVLTGN---GKWVEADPHLRILREVCRSHVRIAEAAATLPPFLGRYRIVLR 273
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 197 TIA---GFDLSSYRDOLTKWKA-----LEVWYACQMEVG---RDKCLPVY 236

QY 274 FEDLAREPLAEIRALYAFTGLTILTPQLEAMHNIHSG-IGKPIEAFTSSRNARVNSQ 332
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 237 YEQLVLRPRSLKRIIDFLGI-----AWSDIVLHEDLIGRP--GGVSLSLKTERTDQ 287

QY 333 A-----WRHALPFTKILRVOEVCAQALQLGYRP 361
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 288 VIKPVNLEALSQWTHIP-RDVRDMAQIAPMLARUGYDP 326

RESULT 12
TPS2_HUMAN
ID TPS2_HUMAN STANDARD; PRT; 377 AA.
AC O60704; O9H0V4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN-TYROSINE SULFOTRANSFERASE 2 (EC 2.8.2.20) (TYROSYLPROTEIN
SULFOTRANSFERASE-2) (TPST-2).
GN TPST2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=98409623; PubMed=9736702;
RA Beisswanger R., Corbeil D., Vannier C., Thiele C., Dohrmann U.,
Kellner R., Ashman K., Niehrs C., Huttner W.B.;
RT "Existence of distinct tyrosylprotein sulfotransferase genes:
molecular characterization of tyrosylprotein sulfotransferase-2.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:11134-11139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98406128; PubMed=9733778;
RA Ouyang Y.-B., Moore K.L.;
RT "Molecular cloning and expression of human and mouse tyrosylprotein
sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in
Caenorhabditis elegans.";
RL J. Biol. Chem. 273:24770-24774(1998).

```

```

[3]
RN SEQUENCE FROM N.A.
RP Bennett E.P.;
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Bridgeman A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RA "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
CC ACIDIC MOTIFS OF POLYPEPTIDES.
CC -!- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + PROTEIN TYROSINE =
CC ADENOSINE 3',5'-BISPHOSPHATE + PROTEIN TYROSINE-O-SULFATE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF061254; AAC34296.1; -
DR EMBL; AF049891; AAC36061.1; -
DR EMBL; A7006198; CAA06906.1; -
DR EMBL; J951115; CAB62950.1; -
DR EMBL; AL136623; CAB66558.1; -
DR MIM; 603126; -
KW Transferase; Transmembrane; Glycoprotein; Signal-anchor;
KW Multigene family.
FT DOMAIN 1 8
FT TRANSMEM 9 25
FT CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT LUMENAL, CATALYTIC (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT O -> L (IN REF. 5).
FT CONFLICT 26 26
FT CONFLICT 73 73
FT CONFLICT 73 73
FT SEQUENCE 377 AA; 41912 MW; A658E50151FDBCL2 CRC64;
Query Match 4.2%; Score 86.5; DB 1; Length 377;
Best Local Similarity 21.3%; Pred. No. 2.4;
Matches 73; Conservative 36; Mismatches 110; Indels 123; Gaps 19;
QY 51 RGSSEFLGQESQPDV-----FYLMPEAWHWTTLSQSAATLHW---AVRDL 96
DB 78 RSGTTLKRAMLDAPHEVCKGEETRIIPVLAHQAW-----SKSGREKLDEAGVLDN 130
QY 97 MRSIFLCMDYVDMPQSRNLSAFFNWNATSRALCSPSPACSAFPFGTTSKO-DVCKTLCT 155
DB 131 -----EVLDA-----AMQAFILEVIAKHEGAPARVLGN 157
QY 156 RQPFSLAREACRSYSHVVLKEVRFNVLQVPLISDPALNKLIVHLVDPRAVLRSREA 215
DB 158 KDPFTLK-----SSVYLSR-----LFP-----NSKFLLMVDRGRASVISMTR 195
QY 216 GPILARDNCIVLGTN----GKWEADPHLRLIREVCRSHVRIAEAAATLKPPFPLRGVRL 271

```

```

DB 196 KVTIA---GFDLSSYRDCLTKWKA---IEVWYAC-----MEVGKEKCLP----- 235
QY 272 YRPEDLAREPLAEALRYAFTGLTLTLPQLEAWTHNTHGSG-IGKPIEAFTHTSSRNARV 330
DB 236 VYIEQVLVHPRRSKLILDFGI-----AWSDAVLHEDLIGKP--GGVSLSKIERST 286
QY 331 SQA-----WRHALPFTKILRVQEVCGAGALQLLGYRP 361
DB 287 DOVIKPVNLEALSQWGTGHP-GDVVRDMAQIAPMLAQLGYDP 327
RESULT 13
RRPL BDV STANDARD; PRT; 1608 AA.
AC P52639;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.46) (LARGE STRUCTURAL PROTEIN)
DE (L PROTEIN) (P180).
CN L.
OS Borna disease virus (BDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales.
OX NCBI_TaxID=12455;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V;
RX MEDLINE=94240137; PubMed=8183914;
RA Briese T., Schneemann A., Lewis A.J., Park Y.-S., Kim S.,
RA Ludwig H., Lipkin W.I.;
RT "Genomic organization of Borna disease virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4362-4366(1994).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U04608; AAA20228.1; -
KW Transferase; RNA-directed RNA polymerase.
KW SEQUENCE 1608 AA; 180254 MW; 68F29C42EFE44D57 CRC64;
Query Match 4.2%; Score 86.5; DB 1; Length 1608;
Best Local Similarity 23.7%; Pred. No. 16; Mismatches 83; Gaps 14;
Matches 64; Conservative
QY 159 FSLAREACRSYSHVVLKEVRFNVLQVL--YPLSDPALNKLIVHLVR-----DPPA 207
DB 1340 YLSWSKSRNOSQVLKAVVRKILCVLYIYPTV-DPAVALDLCHLPALTILVLGGDPAY 1398
QY 208 V--LRREAAGPILAR-----DNGIVLGTNGKWEADP---HLRLIREVCRSHV 251
DB 1399 YERLLEMDLCGAVSSRVDPHSLAGRTHRGFVAGP-----DAGPQVILDRLESVCYAH 1453
QY 252 RIAREATLKPPFPLRGVRLVRFELAREPLAE-----IRALYAF----- 291
DB 1454 CLEE---LEFNAYLDS--ELVDISDMCCUPLATPCALFRFYLSLQSFALMDYSEV 1508
QY 292 -----TGITLTPQLEAWTHNTHGSGI-----GKPIEAFTHTSSRNAR 328
DB 1509 MDLIIMIRGLDIPHLEEFDELIVVGQHLGQVPLVEVYVGVVRKRPVRLARHPWSADLK 1568
QY 329 NVSQAWRHALPFTKILRVQEVCGAGALQLLIG 358
DB 1569 RITVGGRAPCPSAARLR-DEDCQGS-LVIG 1596

```

```

RESULT 14
YPRB_CORGL
ID YPRB_CORGL STANDARD; PRT; 221 AA.
P46584;
DT DT 01-NOV-1995 (Rel. 32, Created)
DT DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL PROTEIN IN PROB 5'REGION (FRAGMENT).
OS Corynebacterium glutamicum (Arevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
NCBI_TaxID=1718;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 17965 / Melassecola;
MEDLIN=96345604; PubMed=8755867;
AnKri S.; Serebrijski I.; Reyes O.; Leblon G.;
"Mutations in the Corynebacterium glutamicum proline biosynthetic
pathway: a natural bypass of the proA step.";
J. Bacteriol. 178:4412-4419(1996).
-- SIMILARITY: SOME, TO B.SUBTILIS OBG C-TERMINAL.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U31230; AAC44173.1; --
InterPro: IPR000765; GPL_OBG.
PROSITE: PS00905; GPL_OBG; PARTIAL.
Hypothetical protein.
NON_TER 1
SEQUENCE 221 AA: 24716 MW.

```

Query Match	4.2%;	Score 85.5;	DB 1;	Length 221;
Best Local Similarity	22.8%;	Pred. NO. 1.5;	Mismatches	53; Conservative 40; Indels 88; Gaps 12;
QY	162	ARACRSYSHVVLKVEYRNFNLGVLYPLL-----SDPALNLRIVHLVRDPRAVLRSR	212	
Db	14	APALKSFAEVL--KYRLKEFGNPFTISAVAKKAUDP-LKYKLEITVDARK-KRPK	69	
QY	213	EAGGPILARDNGIYLGTNGKW-VEADPHLR---LI-----REVCRSHSVRIEAATLKPP	262	
Db	70	EKAESVIIIAKKAVHRTKGQFOIKDPDEVOGGFITTBKPKERWIIQTDFENAEV----	124	
QY	263	PILRGRYRVRFEDLAREPLAIRALYAFTGLTLFPOLAMTHNTHG-----SGIGKPI	317	
Db	125	GYLADRSLKLGTEDLGRAGAHHVGANVTIGLSF-----EWEPMTAGDDPILTGTDV	179	
QY	318	EAFHTSSRNARNVSQAASHALPTFKLIRVQEVCAGALLGYRPYSADOQR	369	
Db	180	RLEQTS----RISAERE-----RASVRRGLIDELYGEDOASRRR	218	

RESULT	15				
ID	ODPA_PIG	STANDARD;	PRT;	389 AA.	
AC	P29804;				
DT	01-APR-1993	(Rel. 25, Created)			
DT	01-APR-1993	(Rel. 25, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	PYRUVATE DEHYDROGENASE E1 COMPONENT ALPHA SUBUNIT, SOMATIC FORM,				
GN	MITOCHONDRIAL PRECURSOR (EC 1.2.4.1) (PDH-E1 TYPE I) (FRAGMENT).				
OR	PDHA1.				
OS	Sus scrofa (Pig).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
[1]
SEQUENCE FROM N.A.
TISSUE=Muscle;
MEDLINE=90370488; PubMed=2395657;
Sermon K., Demeirleir L., Elpers I., Lissens W., Liebaers I.;
"Characterisation of a cDNA for porcine PDH-E1 alpha and comparison
with the human cDNA";
Nucleic Acids Res. 18:4925-4925(1990).
-!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
COPIES OF THREE ENZYMIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
(E3).
-!- CATALYTIC ACTIVITY: PYRUVATE + LIPOAMIDE = S-ACETYL-DIHYDRO-
LIPOAMIDE + CO(2).
-!- COFACTOR: THIAMINE PYROPHOSPHATE.
-!- ENZYME REGULATION: E1 ACTIVITY IS REGULATED BY PHOSPHORYLATION
(INACTIVATION) AND DEPHOSPHORYLATION (ACTIVATION) OF THE ALPHA
SUBUNIT.
-!- SUBUNIT: Tetramer of two alpha and two beta subunits.
-!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

R		EMBL; X52990; CAA37180.1; -.
R	PIR; S20813; DEPGA.	
R	InterPro; IPR001017; El_dh.	
R	Pfam; PF00676; El_dehydrog. 1.	
W	Glycylsis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;	
W	Phosphorylation; Mitochondrion; Transit peptide.	
T	NON_TER	1
I	TRANSIT	<1
T	CHAIN	28
T		29
T		389
T		MITOCHONDRION (BY SIMILARITY).
T		PYRUVATE DEHYDROGENASE E1 COMPONENT
T		ALPHA SUBUNIT, SOMATIC FORM.
T	MOD_RES	231
T		PHOSPHORYLATION (BY SIMILARITY).
T	MOD_RES	292
T		PHOSPHORYLATION (BY SIMILARITY).
T	MOD_RES	299
T		PHOSPHORYLATION (BY SIMILARITY).
Q	SEQUENCE	389 AA; 43121 MW; E9C7DF85389A9A47 CRC64;

```

Query Match      4.2%; Score 85.5; DB 1; Length 389;
Best Local Similarity 17.5%; Pred. No. 3.1;
Matches 47; Conservative 42; Mismatches 89; Indels 91; Gaps 7;

QY 38 GEDRVHVLVLSWRGSGSFLGOLFQH-----PDV FYLMEPAWHVWITLSQGSAA TL 89
db 141 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
184 GKDEV--CLTYIGDGAANQGQIF EAYNKAALWKLPCVFCENRNYGMGTSV ERAAAS TL 240
QY 90 HMAVSLDMSRIFLCDMDVFDAYMPSORNL SAFPFWATSRALCSPAGS AFPRGTIS KQDV 149
db 241 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 YKRGDFTPGLVGDGMDI----- 258

150 KTKLCITGFPSIARACRSYSHVVLKVRFN LQVLP LLSDFALNRYIHLVVDPRAVL 209
db 141 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
259 ---LCVREATF AAAYCRSGKGPILMELQTYRH-----GHSMSDPCVSY 300

210 RSREAAGPIIARDNGIVL---GTNGKWEADPHLRITREVC RSHVKRIASAA TLKP---- 261
db 141 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 RTREIQRVRSKSDPIMLLKDRMVNSNL SVAEE LKEIDVEVRKEIDAAOFATADPEPPL 360
QY 262 -----PPF-LRGRYLVRFDLA 278
db 361 EELGYHYVCNDPFEVRGAGNWKFKKSI S 389

```

us-09-593-828-8.rsp

Fri Feb 1 20:21:12 2002

Search completed: January 31, 2002, 09:51:07
Job time: 325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 09:45:09 ; Search time 42.22 Seconds
(without alignments)
684.239 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRESSKTVIVLLIAQT.....LTLDVLPRGPDHFWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2051	100.0	390	AA72639	Human glycosyl sul
2	1865.5	91.0	418	AA841947	Human ORF1711
3	1729.5	84.3	395	AA72640	Human glycosyl sul
4	1536.5	74.9	395	AA72638	Mouse glycosyl sul
5	1008	49.1	386	AAV39018	Human glycosyl sul
6	970	47.3	386	AA79219	Human transferase
7	958.5	46.7	388	AAV39919	Mouse glycosyl sul
8	665	32.4	483	AAV31656	Mouse N-acetylgluc
9	659.5	32.2	530	AA895367	Human protein sequ
10	656.5	32.0	484	AAV31657	Human N-acetylgluc
11	577.5	28.2	479	AAW52863	Glycosaminoglycan

12	540.5	26.4	458	18	AAW06480	Chick chondroitin
13	513.5	25.0	411	19	AAW61100	Keratan sulphate 6
14	119.5	5.8	596	22	AAV72641	Human glycosyl sul
15	119.5	5.8	1222	22	AAV72642	Human glycosyl sul
16	112.5	5.5	1207	22	AAV72643	Mouse glycosyl sul
17	105	5.1	359	20	AAV06628	C. elegans tyrosyl
18	105	5.1	359	21	AAV06628	C. elegans tyrosyl
19	99.5	4.9	928	22	AAV06627	Drosophila melanog
20	90.5	4.4	380	20	AAV06627	C. elegans tyrosyl
21	90.5	4.4	380	21	AAV06627	A tyrosylprotein s
22	86.5	4.2	330	21	AAV06627	Gene 7 human secre
23	86.5	4.2	330	21	AAV06627	Mouse tyrosylprote
24	86.5	4.2	330	21	AAV06627	A murine tyrosylpr
25	86.5	4.2	376	20	AAV06626	Human tyrosylpro
26	86.5	4.2	377	20	AAV06625	Amino acid sequenc
27	86.5	4.2	377	21	AAV06625	Human zona pelluci
28	86.5	4.2	377	21	AAV06625	Human zona pelluci
29	86.5	4.2	566	15	AAV5207	Human ZPB protein.
30	86.5	4.2	566	15	AAV5207	Human zona pelluci
31	86.5	4.2	566	20	AAV42479	Human oocyte zona
32	86.5	4.2	566	20	AAV42479	Human zona pelluci
33	86.5	4.2	566	21	AAV82214	Borna disease viru
34	86.5	4.2	566	21	AAV52688	Borna disease viru
35	86.5	4.2	566	21	AAV52179	Human polypeptide
36	86.5	4.2	566	21	AAV52983	CHDH. Nocardia sp
37	86.5	4.2	1608	17	AAV98619	Human membrane or
38	86	4.2	1711	17	AAV98605	Corynebacterium gl
39	85.5	4.2	1055	22	AAV39198	C glutamicum prote
40	85.5	4.2	366	12	AAV15475	C. glutamicum acco
41	84.5	4.1	467	22	AAV88483	S. lavendulae Mmer
42	84.5	4.1	484	22	AAV76522	Embryonic stem cel
43	84.5	4.1	491	22	AAV90675	
44	84	4.1	568	21	AAV45789	
45	83.5	4.1	977	18	AAV19258	

ALIGNMENTS

RESULT 1

AAV72639

ID AAV72639 standard; Protein; 390 AA.

AC AAV72639;

XX AAV72639;

DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).

DE Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.

OS Homo sapiens.

XX WO200106015-A1.

PN 25-JAN-2001.

PD 19-JUL-2000; 2000WO-US19741.

XX 20-JUL-1999; 99US-0144694.

PR 13-JUL-2000; 2000US-0593828.

XX (REGC) UNIV CALIFORNIA.

PA

Fri Feb 1 20:21:08 2002

Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnary; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antihypertensive; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergic aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-502362/57.

N-PSDB; AAC76156.

Novel nucleic acids and peptides derived from open reading frame X,

useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

Claim 11: Page 2599-2600; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparkinsonian; nootropic; neuroprotective; antipruritic; anticonvulsant; antiarthritic; immunosuppressant; osteopathic; anticonvulsant; thrombolytic; coagulant; vasotropic; immunostimulant; cardiant; dermatological; immunosuppressive; antidiabetic; hypotensive; antirheumatic; antifungal; antirheumatic; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antihypertensive; antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 418 AA;

Query Match 91.0%; Score 1865.5; DB 21; Length 418;
Best Local Similarity 91.0%; Pred. No. 5e-196;
Matches 356; Conservative 9; Mismatches 25; Indels 1; Gaps 1;

PI Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
DR N-PSDB; AAD02697, AAD02698, AAD02699.
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications -
PT Claim 3; Fig 1; 128pp; English.
PS
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha). GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
XX Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 22; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.9e-216;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPFSSKTVTVLLAQTCLLLFTISRPSPAGGEDRVHVLVLSWRSSSLGQL 60
Db 1 mwlpfsskttvlllaqtcclllfiisrpgpspggedrvhvlvlsrsgssflgql 60
QY 61 FSOHPDVYIMPEAWHVTWTLSSGSAATLHMAYRDLMSRFLCDMDVFDAYMPSQRNLSA 120
Db 61 fsqhpdvfylmepawhvtwtlsggsaatlhmayrldmsrflcdmdvfdaymgsrnlssa 120
QY 121 FFWNATSRALCPSPACSAFPGRTISKQDVCKTLCTRPFSIAREACKRSYSHVVLKEVRF 180
Db 121 ffnwatsralcspacsafprgtiskqdvcktlctrpfsiareacsyshvvlkevrf 180
QY 181 NLQVLPFLSDPALNRIHVLRDPRAVLRSRAAGPILARDNGIVLTGCKWVEADPHL 240
Db 181 nlqvlpflsdpalnrihvldrpravlrsraagpilarndgivltgckwveadphl 240
QY 241 RLREVCRSVHRTAEATIKPPPLRCGRYRLVRFDLAREPLAEIRALYAFGTGLTPQL 300
Db 241 rlrevershrtaeaatikppplrcgrylrvrfdlareplaeiralyaftgtltltpql 300
QY 301 EAWTHNTHSGIGKPTAEHTSSRNARNVSWARHAPLFTKILREVOCAGALQLGYR 360
Db 301 eawhnthsgigkptaehtssrnarnvswarhplftkilevovcagallqlygr 360
QY 361 PVYSADQQRDLTLVLPRGPDHFSWASPD 390
Db 361 pvsadqqrldtlvlprgpdhfswaspd 390

RESULT 2

AAB41947
ID AAB41947 standard; Protein: 418 AA.
XX
AC AAB41947;
XX 08-FEB-2001 (first entry)
DT

QY 1 MWLPFRSSKTVTVLLLAQTCLLLFIISRPSPSPAGGEDRVHVLVLSRWSSGSLGQL 60
 Db 28 mwlpfrssktvtvlllaqtcclllfiisrpgspspaggeardvhlvlsswrsqsfllgql 87
 QY 61 FSOHPDVFYLMPEAWHVTWTLSSGSAATLHMAYRDLMRSLFLCDMDVFDAYM-PQSRNLS 119
 Db 88 fsqhpdvfyimepawhvtwtlsgsaatlhmavrdlmsrlfcdmdvfdaymepgprqs 147
 QY 120 AFFNWATSRALCSPACSAFPRGTISKQDVCKTLCTQPFSLAREACRSYSHVVLKVRP 179
 Db 148 slfgwensralcscapacdliqpdeiiprahcrllscqpfvvevkeacrsyshvvlkevrf 207
 QY 180 FNLQVLPPLSDPALNLRIVHLVRDPRVLRREAAAGPILARDNGIYLGNGKWEADPH 239
 Db 208 fnlqslpplkdslnhlrvldpravlrsreagpilarndngiylgngkweadph 267
 QY 240 LRLIREVCRSHVRIAEATLKPPFLRGVRLVRFEDLAREPLAEIPALYAFGLTLPQ 299
 Db 268 lrlirevcrshvriaeaatlkpplfgrgrylrvfedlareplaeiralyaftgltltpq 327
 QY 300 LEAWITHGSGIGKPIEAFTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQILGY 359
 Db 328 leawithgsgigkpicafhtssrnaarvsgawrhpalpftkilyrvqevcaqalqilgy 387
 QY 360 RPVYSADQQRDLTLDVLPRGPDHFSWSPD 390
 Db 388 rpyvsadqqrldtldvlprgpdhfswaspd 418

RESULT 3
 AAY72640
 ID AAY72640 standard; Protein; 395 AA.
 XX AC AAY72640;
 XX DT
 XX DX
 XX DE
 XX Human glycosyl sulfotransferase-4beta (GST-4beta).
 KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; myasthenia gravis; Sjogren's syndrome; dermatitis;
 KW glomerulonephritis; Grave's disease; hypoparathyroidism; anaemia;
 KW Hashimoto's disease; ulcerative colitis; allergic rhinitis;
 KW myocardiitis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.
 XX OS Homo sapiens.
 XX PN WO200106015-A1.
 XX PD 25-JAN-2001.
 XX PF 19-JUL-2000; 2000WO-US19741.
 XX PR 20-JUL-1999; 99US-0144694.
 XX PR 13-JUL-2000; 2000US-0593828.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Rosen SD, Lee JK, Hemmerich S;
 XX WPI: 2001-138471/14.
 XX DR N-PSDB; AAD02697, AAD02700.
 XX PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications -
 XX Claim 3; Fig 4B; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, Sjogren's syndrome, Hashimoto's
 CC glomerulonephritis, myasthenia gravis, systemic sclerosis, diabetes,
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocardiitis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 XX Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 22; Length 395;
 Best Local Similarity 85.8%; Pred. No. 4e-181;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPFRSSKTVTVLLLAQTCLLLFIISRPSPSPAGGEDRVHVLVLSRWSSGSLGQL 60
 Db 1 mwlpfrsstaavtalllaq-tflllflvsrpgspaggearvhlvlsswrsqsfllgql 59
 QY 61 FSOHPDVFYLMPEAWHVTWTLSSGSAATLHMAYRDLMRSLFLCDMDVFDAYM-PQSRNLSA 120
 Db 60 fndhpdvfyimepawhvtwtlsgsaatlhmavrdlmsrlfcdmdvfdaylpwrnlsd 119
 QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTQPFSLAREACRSYSHVVLKVRPFF 180
 Db 120 lfgwvrsralcspacsaafprgaissavckpicaqsgftlareacrsyshvvlkevrf 179
 QY 181 NLQVLPPLSDPALNLRIVHLVRDPRVLRREAAAGPILARDNGIYLGNGKWEADPHL 240
 Db 180 nlqvlplldspalnrlrvhlvrdrpravlrsreagpilarndngiylgngkweadppl 239
 QY 241 LRLIREVCRSHVRIAEATLKPPFLRGVRLVRFEDLAREPLAEIPALYAFGLTLPQ 300
 Db 240 rrvrevcrshvriaeaatlkpplfgrgrylrvfedlareplaeiralyafglsltpq 299
 QY 301 EAWITHGSGIGKPIEAFTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQILGYR 360
 Db 300 eawithgsgigkpiarfhtssrnaarvsgawrhpalpftkilyrvqevcaqalqilgyr 359
 QY 361 PVSADQQRDLTLDVLPRGPDHFSWAS 388
 Db 360 pvsadqqrldtldvlprgpdhftwas 387

RESULT 4
 AAY72638
 ID AAY72638 standard; Protein; 395 AA.
 XX AC AAY72638;
 XX DT
 XX DX
 XX DE
 XX Human glycosyl sulfotransferase-4 (GST-4).
 KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocardiitis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

KW chromosome 8E1.

XX Mus musculus.

XX WC200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US19741.

XX 20-JUL-1999; 99US-0144594.

XX 13-JUL-2000; 2000US-0593928.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAO02696.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications

XX Claim 3; Fig 2; 128pp; English.

XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4 gene is found on chromosome 8E1.

XX GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation.

XX Sequence 395 AA;

Query Match 74.9%; Score 1536.5; DB 22; Length 395;

Best Local Similarity 76.0%; Pred. No. 6.3e-160;

Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MMLPRFSSTVIVVLLAQITCILLFIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60

DB 1 mrlprfstvmisllmvqtg-ilvflvsrqvpspglgerhvlvlsswsgsvfvgql 59

QY 61 FSGHPDVFYLMPEAWHWTTLTSGSAATLHMAYRDLMSIFLDCMDVFDAYMPQSRNLSA 120

DB 60 fsgdpdvfyimepawhwtlsgsapalhmavrdlmsiflclmdvfdaylpwrnuisd 119

QY 121 FFWNATSRALCSPAPCSAPFRGTIIKQDVCKTLCTQTPFSLAREACRSYSHVVLKEVRFF 180

DB 120 lfqavsvralcspvceafagnisseevckpicafrpfglaqeacssyshvvlkevrff 179

QY 181 NLGVLPPLSDPALNLRIVHLVRDPAVLRSREAGFTIARDNGIVLGTNGKWEADPHL 240

DB 180 nlqvlpllsdpalnrlrvhlvrdravlsrreqtakalardngivlgtngtweadprl 239

QY 241 RLIREVCRSHVIAEAAATLKPPELRGRYRLVRFEDILAREPLAIRALYAFITGLTLFPQL 300

DB 240 rlvnevcrshvriaeaaalhkpplqqrlyrvryediardplrtwrelyafitgltlfpql 299

QY 301 EAWIHTHSGSGIKGPKTEAFTSSRNARNVSVQAWRHALPTFKILRVOEVCAGALQILGYR 360

DB 300 qtwihnthsgsggarreafttsrdalsvsqawrhtlpfakirrvqelcggaqlglyr 359

QY 361 PVIYADQQRDITLDVLPRGPDHFSWAS 388

DB 360 svhseleqrldslldllprgmdsfkwas 387

RESULT 5

AAV39918 standard; Protein; 386 AA.

XX AAV39918;

XX 08-DEC-1999 (first entry)

XX Human glycosyl sulfotransferase-3 protein sequence.

XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing; selectin binding interaction; inflammation; lymphocyte homing; human; secondary lymph organ.

XX Homo sapiens.

XX WO9949018-A1.

XX 30-SEP-1999.

XX 26-FEB-1999; 99WO-US04316.

XX 20-MAR-1998; 98US-0045284.

XX 12-NOV-1998; 98US-0190911.

XX (REGC) UNIV CALIFORNIA.

XX (SYNT) SYNTEX USA INC.

XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;

XX WPI; 1999-580442/49.

XX N-PSDB; AAZ20792.

XX Human and murine glycosyl sulfotransferase 3 and related polynucleotides

XX Claim 2; Fig 1; 59pp; English.

XX This sequence is the human glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.

XX Sequence 386 AA;

Query Match 49.1%; Score 1008; DB 20; Length 386;

Best Local Similarity 52.3%; Pred. No. 6.9e-102;

Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 14 LLIAQITCILLFIISRP-----GPSPPAGGEDRVHVLVLSWRSGSFLGQ 59

DB 1 mlipkknklllflvsqmailalfhmshynisslsmkaqperhmvvlsswsgsvfvgq 60

QY 60 LFSQHPDVFYLMPEAWHWTTLTSGSAATLHMAYRDLMSIFLDCMDVFDAYM-PQSRNL 118

DB 61 lfqdpdvfyimepawhwtlsgsapalhmavrdlmsiflclmdsvfdymepgprq 120

QY 119 SAFTFNATSRALCSPAPCSAPFRGTIIKQDVCKTLCTQTPFSLAREACRSYSHVVLKEVR 178

DB 121 sslfqwensralcspacdiipqdeilpbrahrilcsqgpfvevkeacrsyshvvlkev 180

QY 179 FFNLQVLPPLSDPALNLRIVHLVRDPAVLRSREAGFTIARDNGIVLGTN-GKWEAD 237


```

Db 181 ffnlqslpdkpslnhvlrvdravfrsrtkqgldmsrvmgqheqklkdd 240
QY 238 PHRLIREVCRSHVRIAPAAALKP-PPFLRGYRVLVREEDLAREPLAETRALYATGLTL 296
Db 241 qpyyvmqvicqsleiyk--tiqslpkalqeryllrvyedarapvqtsrmyefvgief 298
QY 297 TPQLEAWIHNTHGSGICKPIEAFTSSRNARNVSNQAWRHALPFTKILRVQEVCAQALQ 356
Db 299 lphlqtwvhnitrgkmgd--hafthnrdalnvsgawrslpyekvsrlkacgdaml 356
QY 357 LGYRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 lgyrhviseqegrnllldl-----stwtvpe 383

```

RESULT 6

AA79219

XX AAY79219 standard; Protein; 386 AA.

AC AAY79219;

XX AAY79219;

DT 19-JUN-2000 (first entry)

XX Human transferase TRNSFS-11.

DE DE

KW Transferase; TRNSFS-11; human; antitumour; cell proliferation;
 KW gastrointestinal disorder; developmental disorder;
 KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle disorder; immunological disorder; inflammation;
 KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.

XX Homo sapiens.

XX

FH Key Location/Qualifiers

```

FT Modified-site 121
FT /note= "potential O-phosphorylation"
FT Modified-site 107
FT /note= "potential O-phosphorylation"
FT Modified-site 217
FT /note= "potential O-phosphorylation"
FT Modified-site 252
FT /note= "potential O-phosphorylation"
FT Modified-site 364
FT /note= "potential O-phosphorylation"
FT Modified-site 380
FT /note= "potential O-phosphorylation"
FT Modified-site 35
FT /note= "potential O-phosphorylation"
FT Modified-site 50
FT /note= "potential O-phosphorylation"
FT Modified-site 81
FT /note= "potential O-phosphorylation"
FT Modified-site 287
FT /note= "potential O-phosphorylation"
FT Modified-site 243
FT /note= "potential O-phosphorylation"
FT Modified-site 30
FT /note= "potential O-phosphorylation"
FT Modified-site 308
FT /note= "potential N-glycosylation"
FT Modified-site 329
FT /note= "potential N-glycosylation"
FT Domain 7..23
FT /note= "potential N-glycosylation"
FT /note= "transmembrane domain"

```

XX WO200014251-A2.

XX 16-MAR-2000.

XX 09-SEP-1999; 99WO-US20989.

XX 10-SEP-1998; 98US-0150657.

PR

```

PR 04-NOV-1998; 98US-0186779.
PR 11-MAY-1999; 99US-0133642.
XX (INCY-) INCYTE PHARM INC.
PI Tang YF, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
PI Hillman JL, Azimzai Y;
XX WPI; 2000-256996/22.
DR N-PSDB; AAZ94211.
XX Human transferase proteins useful for preventing, diagnosing and
PT treating cancers and developmental, gastrointestinal, genetic,
PT immunological, neurological, reproductive and smooth muscle disorders -
XX Claim 1; Page 90-91; 113pp; English.
CC The present sequence is that of human transferase TRNSFS-11, 1 of
CC 15 claimed novel human transferase proteins of the invention (see
CC AAY79209-23). The sequence was deduced from a cDNA clone (see
CC AAZ94211) isolated from a gallbladder library. It shows homology to
CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
CC expressed in dermatologic and gastrointestinal tissues,
CC especially those associated with inflammation and cell
CC proliferation. The new human transferases and polynucleotides can
CC be used in the diagnosis, prevention and treatment of cancer,
CC developmental disorders, gastrointestinal disorders, genetic
CC disorders, immunological disorders, neurological disorders,
CC reproductive disorders, and smooth muscle disorders. The
CC polypeptides can also be used to raise antibodies, and to screen
CC for agonists and antagonists of transferase activity.
XX Sequence 386 AA;
SQ

```

Query Match

47.3%; Score 970; DB 21; Length 386;

Best Local Similarity 51.3%; Pred. No. 1e-97;

Matches 202; Conservative 52; Mismatches 112; Indels 28; Gaps 7;

```

QY 14 LLLAQITCILLFIISRP-----GPSSPAGGEDRVHVLVLSWSRSGSFLQG 59
Db 1 mlpdkmklfivsqmailalfhmyshmisslsmkagpermbvlvsswsgsfvg 60
QY 60 LFSQHPDVFYLMPEAWHVVWTLTSGSAATLHMAVRDLMSIFLCLDMVDYAYM-PQSRNL 118
Db 61 lfgqdpdvfyimepawhvmvtfkqstawlhmavrdliravflcdmsvfdamypgrrq 120
QY 119 SAFFWATSRALCSPPACSAFFRGITISKQDVCKTLCTROPFSILARACKRSYHVLKEVR 178
Db 121 ssifqwnsralscapacdilpqdesspgltagscavnsplklekacrsyshvlkev 180
QY 179 FENLOVLYPLLSDPALNLRIVHLVRDPRAVLRSRAGPILARDNGIVLGTN-GKWVEAD 237
Db 181 ffnlqslpdkpslnhvlrvdravfrsrtkqgldmsrvmgqheqklkdd 240
QY 238 PHRLIREVCRSHVRIAPAAALKP-PPFLRGYRVLVREEDLAREPLAETRALYATGLTL 296
Db 241 qpyyvmqvicqsleiyk--tiqslpkalqeryllrvyedarapvqtsrmyefvgief 298
QY 297 TPQLEAWIHNTHGSGICKPIEAFTSSRNARNVSNQAWRHALPFTKILRVQEVCAQALQ 356
Db 299 lphlqtwvhnitrgkmgd--hafthnrdalnvsgawrslpyekvsrlkacgdaml 356
QY 357 LGYRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 lgyrhviseqegrnllldl-----stwtvpe 383

```

RESULT 7

AAY39919

ID AAY39919 standard; Protein; 388 AA.

XX

AC AAY39919;

XX 08-DEC-1999 (first entry)
 DT Mouse glycosyl sulfotransferase-3 protein sequence.
 XX
 DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
 KW secondary lymph organ.
 XX
 XX Mus sp.
 OS
 XX W09949018-AL.
 PN
 XX 30-SEP-1999.
 PD
 XX 26-FEB-1999; 99WO-US04316.
 PF
 XX 20-MAR-1998; 98US-0045284.
 XX 12-NOV-1998; 98US-0190911.
 PR
 XX (REGC) UNIV CALIFORNIA.
 XX (SYNT) SYNTAX USA INC.
 PA
 XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
 PI
 XX WPI; 1999-580442/49.
 DR
 XX N-PSDB; AAZ20793.
 DR
 XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides -
 PT
 XX Claim 2; Fig 4; 59pp; English.
 PS
 XX This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions.
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 XX
 XX Sequence 388 AA;
 SQ

Query Match 46.7%; Score 958.5; DB 20; Length 388;
 Best Local Similarity 55.6%; Pred. No. 1.9e-96;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;
 QY 42 VHVILSSWRSGSFLGOLFQSDHPVFLMEPAHWVHTLISQGSAAATLHMAVRDLMSIF 101
 DB 42 VHVILSSWRSGSFLGOLFQSDHPVFLMEPAHWVHTLISQGSAAATLHMAVRDLMSIF 101
 QY 102 LQDMVFDATM-PQSRNLISAFENWATSRALCSPACSAFPRGTISKODVCKTLCTROPFS 160
 DB 102 LQDMVFDATM-PQSRNLISAFENWATSRALCSPACSAFPRGTISKODVCKTLCTROPFS 160
 QY 102 LQDMVFDATM-PQSRNLISAFENWATSRALCSPACSAFPRGTISKODVCKTLCTROPFS 161
 DB 102 LQDMVFDATM-PQSRNLISAFENWATSRALCSPACSAFPRGTISKODVCKTLCTROPFS 161
 QY 161 LAREACRSYSHVVLKEVRFNNQVLYPLSDPALNLRVHIVDPRAVLSREAGPILA 220
 DB 161 LAREACRSYSHVVLKEVRFNNQVLYPLSDPALNLRVHIVDPRAVLSREAGPILA 220
 QY 162 MVEKACSHGFVFLKEVRFNNQVLYPLSDPALNLRVHIVDPRAVLSREAGPILA 221
 DB 162 MVEKACSHGFVFLKEVRFNNQVLYPLSDPALNLRVHIVDPRAVLSREAGPILA 221
 QY 221 RDNGIVLGTNGKWWADPHRLKIRE-----VCRSHVRIAEATLKPFPFLRGY 269
 DB 221 RDNGIVLGTNGKWWADPHRLKIRE-----VCRSHVRIAEATLKPFPFLRGY 269
 QY 222 VDSHIVIG-----qhlteikeedqpyyamklicksqrdivkaqtl--pealqrry 270
 DB 222 VDSHIVIG-----qhlteikeedqpyyamklicksqrdivkaqtl--pealqrry 270
 QY 270 RLVPFEDLAREPLAIFRLAFTGLTLPQLEAMHTNTHGSGICKPKEAFHTSSRNARN 329
 DB 270 RLVPFEDLAREPLAIFRLAFTGLTLPQLEAMHTNTHGSGICKPKEAFHTSSRNARN 329
 QY 271 Lfryedlvraplaqttrlykfvglfdlphlgtwvynvtrkgmgq--hafntnarnaln 328
 DB 271 Lfryedlvraplaqttrlykfvglfdlphlgtwvynvtrkgmgq--hafntnarnaln 328
 QY 330 VQSAWRHALPTFKILRVQEVGACAGALQILGYRPVYSADQQRDLTLDLV 376
 DB 330 VQSAWRHALPTFKILRVQEVGACAGALQILGYRPVYSADQQRDLTLDLV 376
 QY 329 VQSAWRHALPTFKILRVQEVGACAGALQILGYRPVYSADQQRDLTLDLV 375
 DB 329 VQSAWRHALPTFKILRVQEVGACAGALQILGYRPVYSADQQRDLTLDLV 375

RESULT 8
 AAY31656
 ID RAY31656 standard; Protein; 483 AA.
 XX
 AC AAY31656;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 XX Mouse N-acetylglucosamine-6-O-sulfotransferase.
 DE
 XX N-acetylglucosamine-6-O-sulfotransferase; mouse; GlyCAM-1;
 KW L-selectin ligand.
 KW
 XX Mus musculus.
 OS
 XX EP943688-A2.
 PN
 XX 22-SEP-1999.
 PD
 XX 04-MAR-1999; 99EP-0301530.
 PF
 XX 24-JUN-1998; 98JP-0177844.
 PR
 XX 05-MAR-1998; 98JP-0054007.
 PR
 XX (SEK) SEIKAGAKU CORP.
 PA
 XX Habuchi O, Kadomatsu K, Kannagi R, Muramatsu H;
 PI Muramatsu T, Uchimura K;
 PI
 XX WPI; 1999-520337/44.
 DR
 XX N-PSDB; AAX87820.
 DR
 XX New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful
 PT for synthesis of sugar chains, e.g. GlyCAM-1
 PT
 XX Claim 1; Page 24-25; 41pp; English.
 PS
 XX The present sequence represents mouse N-acetylglucosamine-6-O-
 CC sulfotransferase, an enzyme capable of transferring a sulfate group
 CC from a sulfate group donor to a hydroxyl group at the 6 position of
 CC an N-acetylglucosamine residue located at the non-reducing end of
 CC an oligosaccharide of formula GlcNAc β Gal β 1-4GlcNAc, where
 CC GlcNAc = N-acetylglucosamine residue, Gal = galactose residue,
 CC β 1-3 = β 1-3 glycosidic linkage, and β 1-4 = β 1-4
 CC glycosidic linkage. The enzyme is useful for the synthesis of
 CC sugar chains such as GlyCAM-1, a ligand of L-selectin that is
 CC involved in homing of lymphocytes and rolling of leukocytes
 CC occurring at the early stage of inflammation. DNA encoding the
 CC enzyme (see AAX87820) is expected to be used for the large-scale
 CC production of N-acetylglucosamine-6-O-sulfotransferase, or
 CC artificial synthesis of GlyCAM-1 using transformants which harbour
 CC the DNA.
 XX
 XX Sequence 483 AA;
 SQ

Query Match 32.4%; Score 665; DB 20; Length 483;
 Best Local Similarity 40.4%; Pred. No. 4.5e-64;
 Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;
 QY 37 GGEDRVHVLVLSNRSGSFLGOLFQSDHPVFLMEPAHWVHTLISQGSAAATLHMAVRDL 96
 DB 37 GGEDRVHVLVLSNRSGSFLGOLFQSDHPVFLMEPAHWVHTLISQGSAAATLHMAVRDL 96
 QY 97 MRSIFLCMDVFDATMPO---SRNLS--AFENWATSRALCSPACSAFPRGTISKOD--V 149
 DB 97 MRSIFLCMDVFDATMPO---SRNLS--AFENWATSRALCSPACSAFPRGTISKOD--V 149
 QY 173 lsalyrdlvraplaqttrlykfvglfdlphlgtwvynvtrkgmgq--hafntnarnaln 232
 DB 173 lsalyrdlvraplaqttrlykfvglfdlphlgtwvynvtrkgmgq--hafntnarnaln 232
 QY 150 CKTLCTROPFSILAREACRSYSHVVLKEVRFNNQVLYPLSDPALNLRVHIVDPRAVL 209
 DB 150 CKTLCTROPFSILAREACRSYSHVVLKEVRFNNQVLYPLSDPALNLRVHIVDPRAVL 209
 QY 233 ckk-cppqr-larfeecrkyrtvvlkgrvfdvavlapllkdpaldkvlhivrdprava 291
 DB 233 ckk-cppqr-larfeecrkyrtvvlkgrvfdvavlapllkdpaldkvlhivrdprava 291

QY 210 RS-----REAGPILARD-----NGIVLGT--NGKWEADPH-IRLIREV 246
Db 292 szsirrhgllreslqvvrspgrahrmpflleaagbkigakkeggpgpadyhalgamevi 351
QY 247 CRSHVRIAEAAATLKPPPTLGRYRIVREFDLAREPLAEIRALYAFITGLTLTPQLEAWITHN 306
Db 352 csmaktiqta-lgppdwlqghylvryedlvdpvklrrydfvgllvspemeqfalm 410
QY 307 ITHGSG-IGKPLEAPWTSRRNARNVSQARWALPPTKILRVOEVCAGALQLLGYRPYSA 365
Db 411 mtsgsgssskp--fvssaratqaanwrtalfqgikveefcygpmavlgvrvnsp 467
QY 366 DQORLITLDLV 376
Db 468 eevkdlstll 478

RESULT 9
AAB95367
ID AAB95367 standard; Protein; 530 AA.
AC AAB95367;
XX
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:17679.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17679; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by the


```

XX AC AAW06480;
XX
XX DT 25-FEB-1997 (first entry)
XX
XX DE Chick chondroitin 6-sulphotransferase.
XX
XX KW Chondroitin 6-sulphotransferase; C6ST; chondroitin sulphate;
XX proteoglycan.
XX OS Gallus sp.
XX
XX FH Key
XX FT Peptide
XX FT /label= Sig_peptide
XX FT /label= 34..458
XX FT /note= Mat_protein
XX FT /note= "preferred polypeptide (Claim 5)"
XX FT /note= "preferred polypeptide (Claim 3)"
XX FT /note= "preferred polypeptide (Claim 4)"
XX FT /note= "preferred polypeptide (Claim 6)"
XX FT /label= 24..37
XX FT /label= Transmembrane_domain
XX FT /label= 62..64
XX FT /label= Glycosylation
XX FT /note= "potential N-glycosylation site"
XX FT /label= 73..75
XX FT /label= Glycosylation
XX FT /note= "potential N-glycosylation site"
XX FT /label= 95..97
XX FT /label= Glycosylation
XX FT /note= "potential N-glycosylation site"
XX FT /label= 236..238
XX FT /label= Glycosylation
XX FT /note= "potential N-glycosylation site"
XX FT /label= 399..401
XX FT /label= Glycosylation
XX FT /note= "potential N-glycosylation site"
XX FT /label= 443..445
XX FT /label= Glycosylation
XX FT /note= "potential N-glycosylation site"
XX
XX EP745668-A2.
XX
XX PD 04-DEC-1996.
XX
XX PF 30-MAY-1996; 96EP-0303887.
XX
XX PR 31-MAY-1995; 95JP-0134358.
XX
XX PA (SEIK ) SEIKAGAKU KOGYO CO LTD.
XX
XX FI Fukuta M, Habuchi O;
XX
XX DR WPI; 1997-013692/02.
XX
XX DR N-PSDB; AAT45037.
XX
XX DNA encoding chondroitin 6-sulpho:transferase - for recombinant
XX production of C6ST, for use in industrial processes
XX
XX PS Claim 2; Page 18-20; 30pp; English.
XX
XX CC Chick embryo chondrocyte chondroitin 6-sulphotransferase (C6ST)
XX (AAW06480) is capable of catalysing the transfer of a sulphate group
XX from 3'-phosphadenosine 5' phosphosulphate to the hydroxyl group
XX at the C-6 position of the N-acetylglucosamine residue of
XX chondroitin. Its amino acid sequence was deduced from a cDNA clone
XX (AAT45037) isolated from a chick embryo chondrocyte cDNA library.
XX Isolation of the cDNA allows the mass prodn. of C6ST in transformed
XX host cells. Recombinant C6ST polypeptides are useful for analysing

```

```

CC the activities of chondroitin sulphate and for modifying its
CC function, and for raising antibodies.
XX
XX SQ Sequence 458 AA;
XX
XX Query Match 26.4%; Score 540.5; DB 18; Length 458;
XX Best Local Similarity 34.3%; Pred. No. 2e-50;
XX Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;
XX
XX QY 36 AGGEDRVHVLSSWRSGSFLGQLFSQHPDVYFLMEPAWHV--WTTLSQGSAAITHMAV 93
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 107 aaeprrhvlmatrtgssvfgefnqgnifylfeplwhiertvttepgganavgsal 166
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 94 --RDLMSIFLCLMDVFDAYM---PQSRNLSAFFENKATSRALCSPACSAFPRGTISKOD 148
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 167 vrgdvlqlllclldlyilesfispapeehltaalfrgsshsiceepvtpeikvkfeyh 226
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 149 VKTILCTROPFLSLAREACRSYSHVVLEKVRFFNLQVLYPLLSDPALNLRIVHLVRDPRAV 208
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 227 cknrrcgplnitlaeacrrkqhmalktvrirqlleflqlaepdrldiriqlvrdprav 286
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 209 LRSREAAGPIIARDNGIVLGTNGKWE-----ADPHRLIREVCRSHVRAEAAIL 259
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 287 lvsmva-----fsgkyeswkkwaaegeapldqdevgrlrgnces-irlsaelgl 335
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 260 KPFPFLRGVRLVRFEDLAREPLAEIRALYAFGLTLPQLEAWTHNTHSGIGKPKTEA 319
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 336 rqrwrlrgrymlvryedvaraplkalemyrfagihptpgveewlrantgap--qdsng 392
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 320 FHVSSRNARNVQAWRHALPFTKILRVQVCAGALQLLIGYRPVYSADQQRDLTLDLVLP 379
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 393 iystqksseqfekwrfspfklaqvvgdacepamrlfgyklassagclnrsisl-lee 451
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX QY 380 GP 381
XX | |
XX Db 452 gp 453
XX
XX RESULT 13
XX AAW61100
XX ID AAW61100 standard; Protein; 411 AA.
XX
XX AC AAW61100;
XX
XX DT 14-SEP-1998 (first entry)
XX
XX DE Keratan sulphate 6-sulphotransferase.
XX
XX KW KSGal6ST; keratan sulphate 6-sulphotransferase;
XX KW chick chondroitin 6-sulphotransferase; C6ST; phage Lambda; hybridization;
XX KW expression vector; COS-7 cells; Bluescript plasmid; galactose;
XX KW keratan sulphate; chondroitin.
XX
XX OS Homo sapiens.
XX
XX PN EP845533-A2.
XX
XX PD 03-JUN-1998.
XX
XX PF 27-NOV-1997; 97EP-0309564.
XX
XX PR 29-NOV-1996; 96JP-0320535.
XX
XX (SEIK ) SEIKAGAKU KOGYO CO LTD.
XX
XX FI Fukuta M, Habuchi O;
XX
XX DR WPI; 1998-286750/26.
XX
XX DR N-PSDB; AAV36418.
XX
XX Keratan sulphate 6-sulpho-transferase polypeptide - transfers
XX sulphate from sulphate donor to galactose 6-hydroxy group etc.

```

KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis
XX	asthma; hypersensitivity; rheumatic fever; tissue rejection.
XX	
XX	Homo sapiens.
XX	
PN	WO200106015-A1.
XX	
XX	25-JAN-2001.
PD	
XX	19-JUL-2000; 2000WO-US19741.
XX	
XX	20-JUL-1999; 980US-0144694.
PR	13-JUL-2000; 2000US-0593828.
PR	
XX	(REGC) UNIV CALIFORNIA.
PA	
XX	Rosen SD, Lee JK, Hemmerich S;
XX	
PI	
XX	WPI; 2001-138471/14.
XX	N-PSDB; AAD02702, AAD02703.
DR	
DR	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX	and theraeutic agent screening applications -
XX	
PT	

Claim 3; Fig 3; 120pp; English.

The present sequence is human glycosyl sulfotransferase-6 (GST-6) fragment.

GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation.

systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation.

```

XX      Sequence      596 AA;
SQ
      Query Match      5.8%; Score 119.5; DB 22; Length 596;
      Best Local Similarity 21.6%; Pred. No. 0.00052;
      Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;
      28 SRFGPSSPAGGDRHVHVLVLSRRSGSSFLGQLFSOHDPDVFYLMPEFAWHVWTTLSGSGAA 87
      223 skkslsseghmdlpdvatslpgsgaelklqfnssdflrvp-----ta 270
      88 TLHMAVRDLMRSTEL-CMDVFDAYMPSRNLSAFFNWAISRA-----LCSP 134
      271 yidpeteleidsfvadacewksdi---rsghflrlrgwlqslvgdtklhlqnihlep- 326
      135 ACSAFPRTGTSK-----QDVCKTLCTRPFLSLAREACRSYSHVVLKEVFENLQVLYPL 189
      327 -----nrgklagfammkdkkrfkreslpeqsgmqgafdrdaeylralrhvvyyps 381
      190 SDFALNL-----RVHLVYDRPFAVLRSR-EAAGPLIARDNGI--- 225
      382 arpvlslssgswtlklhffdevgalmsralryvdrpawlysmlynkspslysknvpeh 441
      226 -----VLGTNGKW-----VEADPHLRLLREVCR-----SHVRIAE-AATIK-P 261
      442 laklfkieggkgknlsngyafeyep---lrkelsksksnavslshlwantaalarin 498
      262 PPFLRGYRYVRFPDLAREPLAETRALYAFQTGLTLTQLEAWTHNTHGSGIKGPTEAFH 321
      190 SDFALNL-----RVHLVYDRPFAVLRSR-EAAGPLIARDNGI--- 225
      382 arpvlslssgswtlklhffdevgalmsralryvdrpawlysmlynkspslysknvpeh 441
      226 -----VLGTNGKW-----VEADPHLRLLREVCR-----SHVRIAE-AATIK-P 261
      442 laklfkieggkgknlsngyafeyep---lrkelsksksnavslshlwantaalarin 498
      262 PPFLRGYRYVRFPDLAREPLAETRALYAFQTGLTLTQLEAWTHNTHGSGIKGPTEAFH 321

```

Db 499 tdlptsyqlvkfedivhfpqkterifafliplspaslnqilfatstnlfylvpyeg-e 557

QY 322 TSSRNARNVSOAQRHLPFTKILRVQEVOCAGALQLGY 359

Db 558 isptn-----tnwkgnlprdeiklienicwtlmdrlgy 591

RESULT 15

AAV72642

ID AAV72642 standard; Protein; 1222 AA.

AC AAV72642;

XX 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-6 (GST-6).

XX Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive;

KW therapy; selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenomatosis;

KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

KW asthma; hypersensitivity; rheumatic fever; tissue rejection.

OS Homo sapiens.

XX

XX

FH Key Location/Qualifiers

FT Domain 851..1223

FT /label= C-terminal_sulfotransferase_domain

XX WQ0200106015-A1.

XX

PD 25-JAN-2001.

XX

PF 19-JUL-2000; 2000WO-US19741.

XX

PR 20-JUL-1999; 990US-0144594.

PR 13-JUL-2000; 2000US-0593828.

XX

XX (REGC) UNIV CALIFORNIA.

XX

XX Rosen SD, Lee JK, Hemmerich S;

PI

XX WPI; 2001-138471/14.

DR N-PSDB; AAD02702, AAD02704.

XX

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

PT diagnostic and therapeutic agent screening applications -

XX

PS Claim 3; Fig 5B; 128pp; English.

XX

XX The present sequence is human glycosyl sulfotransferase-6 (GST-6).

XX GST is a type 2 membrane protein useful for inhibiting a binding event

XX between a selectin and a selectin ligand, which comprises contacting the

XX selectin with a non-sulphated selectin ligand, GST and a small molecular

XX agent that inhibits the sulphation activity of GST. GST is also useful

XX in inhibiting a selectin mediated binding event. GST is useful in gene

XX therapy to treat disorders such as acute or chronic inflammation,

XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis

XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,

XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's

XX disease, Grave's disease, adenomatosis, hypoparathyroidism, pernicious

XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,

XX dermatitis, myocarditis, regional enteritis, adult respiratory distress

XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,

XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection

XX during transplantation.

XX

SQ Sequence 1222 AA;

Query Match 5.8%; Score 119.5; DB 22; Length 1222;

Best Local Similarity 21.6%; Pred. No. 0.0015;

Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY 28 SRPGSPSPAGGEDRVHVLVLSWRSGSSTFGQLFSQHDPDVFYIMEPAWHVMTWTLISQGSAA 87

Db 849 skxslsssegthmdlpdvvitslpgsgaeilkqlfssdflyirvp-----ta 896

QY 88 TLHMAVRDLMRSLFL--CDMDVFDAYMPOSRLNLSAFNNATSR-----LCSP 134

Db 897 yidipeteleidsfdadcekwksdi---zsgthfrllrgwlqslvqgdklhlgnihhep- 952

QY 135 ACSAPFRGTISK-----QDVCKTLCTRPQPSLAREACRSYSHVVLKEVRFNLOVLYPLL 189

Db 953 -----nrgklagyfamnkdkrkfkreslpegrsqmkgafrdaeyfialrrhlvyyps 1007

QY 190 SPALNL-----RIVHLVRDPRAVLRSR-FAAGPTILARDNGI--- 225

Db 1008 arpvslslsgswtlklhffgevlgsamrallyivrdprawyismlynskpslysknvpeh 1067

QY 226 -----VLGTNGKW-----VEADPHLRLIREVCR-----SHVRIAP-AAATLK-P 261

Db 1068 laklfkiegkgkcnlnsgyafeyep---lrkelsksksnavsllshlwlantaaalrin 1124

QY 262 PPFLRGYRLVRFEDLAREPLAEIRALYAFGLTLPQLEAWIHNIHSGSGIGKPIEAPH 321

Db 1125 tdlptsyqlvkfedivhfpqkterifafliplspaslnqilfatstnlfylvpyeg-e 1183

QY 322 TSSRNARNVSOAQRHLPFTKILRVQEVOCAGALQLGY 359

Db 1184 isptn-----tnwkgnlprdeiklienicwtlmdrlgy 1217

Search completed: January 31, 2002, 09:45:11

Job time: 219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 18:18:06 ; Search time 315.85 seconds
(without alignments)
3183.924 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtgctgcacggtctc.....gctggcgcgcctgactga 1173

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1173	100.0	2044	22	Human glycosyl sul
2	1173	100.0	2170	22	Human glycosyl sul
3	1173	100.0	160552	22	Human glycosyl sul
4	1027.2	87.6	2988	22	Human ORFX ORF1711
5	924.6	78.8	1694	22	Human glycosyl sul
6	745.2	63.5	1969	22	Human glycosyl sul
7	400	34.1	2032	20	Human glycosyl sul
8	377	32.1	2065	21	Human transferase
9	354	30.2	1926	20	Mouse glycosyl sul
10	235.2	20.1	2156	19	Glycosaminoglycan
11	196.8	16.8	2354	18	Chick chondroitin

12	153.8	13.1	3029	22	AAH17922	Human cDNA sequenc
13	148.8	12.7	2409	20	AAH87821	Human N-acetylgluc
14	139	11.8	1458	19	AAV36418	Keratan sulphate 6
15	133	11.3	2150	20	AAH87820	Mouse N-acetylgluc
16	64.6	5.5	43280	18	AAH80413	Tylosine synthase
17	57.8	4.9	731	20	AAZ24563	Human lung tumor a
18	57.8	4.9	731	21	AAZ24563	Human lung cancer-
19	57.2	4.9	30001	18	AAZ65802	Total DNA sequence
20	57.2	4.9	30001	20	AAZ65802	S. aureofaciens DN
21	55	4.7	114955	20	AAH53491	Human adenosine A1
22	54.8	4.7	58857	21	AAH58471	Nucleotide sequenc
23	54.6	4.7	4689	21	AAZ87299	S. venezuelae macr
24	54.6	4.7	36778	21	AAZ87318	S. venezuelae pik
25	54.6	4.7	37948	21	AAZ87285	S. venezuelae pik
26	54.6	4.7	38506	21	AAZ75633	Nucleotide sequenc
27	54.6	4.7	38506	21	AAZ75633	Recombinant cosmid
28	54.6	4.7	38506	21	AAZ58001	FLGA insert stabl
29	54	4.6	795	19	AAV55830	Micromonospora DNA
30	53.8	4.6	109519	22	AAH58093	Epstein Barr Virus
31	53.4	4.6	1925	20	AAH90924	Amycolatopsis medi
32	53.2	4.5	53789	19	AAV21187	Pseudomonas fluore
33	52.8	4.5	1683	22	AAH03820	Epstein Barr virus
34	52.8	4.5	1926	21	AAH50254	EBV tethering prot
35	52.8	4.5	1926	22	AAH82902	Nucleotide sequenc
36	52.8	4.5	2580	21	AAH75454	Anti-sense strand
37	52.8	4.5	5452	20	AAH90923	Vector pshuttle DN
38	52.8	4.5	8705	20	AAZ23778	Vector plasmid pCM
39	52.8	4.5	9600	19	AAV21683	Nucleotide sequenc
40	52.8	4.5	10380	20	AAZ22248	Plasmid pCISBON f
41	52.8	4.5	10596	14	AAH51731	Plasmid pCISBON f
42	52.8	4.5	10596	17	AAH40348	Nucleotide sequenc
43	52.8	4.5	10596	20	AAH15650	DNA clone pCFK C1
44	52.8	4.5	16080	21	AAH59553	Swine pseudorabies
45	52	4.4	1734	21	AAH40350	Pseudorabies virus
			4897	11	AAQ03259	

ALIGNMENTS

RESULT 1	
AD02699	AD02699 standard; cDNA: 2044 BP.
XX	XX
AC	AC
XX	XX
DT	02-MAY-2001 (first entry)
XX	XX
DE	Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.
XX	XX
KW	Human; glycosyl sulfotransferase-4alpha: GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW	demylinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW	chromosome 16q23.1; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PH	Key
FT	Location/Qualifiers
FT	218..1390
FT	/tag= a
FT	/product= "Human glycosyl sulfotransferase-4alpha
FT	(GST-4alpha)"
FT	/note= "CDS is specifically claimed as SEQ ID NO: 4
FT	in claim 6 (page no: 41) of the specification"
XX	XX
PN	WC0200106015-A1.
XX	XX
PD	25-JAN-2001.

XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX FT 5'UTR 9..343
 XX FT CDS /*tag= a
 344..1516
 FT /*tag= b
 FT /product= "Human glycosyl sulfotransferase-4alpha
 (GST-4alpha)"
 FT /note= "CDS is specifically claimed as SEQ ID NO: 4
 in claim 6 (page no: 41) of the specification"
 FT 3'UTR 1517..2134
 FT /*tag= c
 XX WO200106015-A1.
 XX PN 25-JAN-2001.
 XX PD 19-JUL-2000; 2000WO-US19741.
 XX PF 20-JUL-1999; 99US-0144694.
 XX PR 13-JUL-2000; 2000US-0593825.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Rosen SD, Lee JK, Hemmerich S;
 XX PT WPI: 2001-138471/14.
 XX PS P-PSDB; AAV72639.
 XX PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 diagnostic and therapeutic agent screening applications -
 XX PS Claim 6; Page 62; 128pp; English.
 XX CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
 alpha) cDNA. GST-4 gene is found on Chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 XX CC Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;

Query Match 100.0%; Score 1173; DB 22; Length 2170;
 Best Local Similarity 100.0%; Pred. No. 2.5e-217;
 Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 0;
 QY 1 atgtggctgcacggtctccagacagacagtgacagtgctctctctgacacaccacc 60
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 tgcctctgct 120
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 404 tgcctctgct 463
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 121 cgtgtgacagtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgt 180
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 464 cgtgtgacagtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgt 523
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 181 ttacgacagacaccccgacgttcttctacatgagagcccgctgagcatgtgtgagacacc 240
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 524 ttacgacagacaccccgacgttcttctacatgagagcccgctgagcatgtgtgagacacc 583
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 241 ctgtgcagagcagc 300
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 584 ctgtgcagagcagc 643
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 301 tttttgctgacatggacgtttttgctgacatgacatgacatgacatgacatgacatgac 360
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 644 tttttgctgacatggacgtttttgctgacatgacatgacatgacatgacatgacatgac 703
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 361 tttttcaactggcaacagc 420
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 704 tttttcaactggcaacagc 763
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 421 cgagcacacatcagcaacagc 480
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 764 cgagcacacatcagcaacagc 823
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 481 ctggcccgagagc 540
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 824 ctggcccgagagc 883
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 541 aacctgcagaggtgctctacacccgctctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 884 aacctgcagaggtgctctacacccgctctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 943
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 601 ctggctgc 660
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 944 ctggctgc 1003
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 661 cgcgcacacgc 720
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1004 cgcgcacacgc 1063
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 721 cgcctgattcgc 780
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1064 cgcctgattcgc 1123
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 781 cgcgcacaccttctcgc 840
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1124 cgcgcacaccttctcgc 1183
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 841 cgcctgagagatcgc 900
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1184 cgcctgagagatcgc 1243
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 901 gaggcctgattcgc 960
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1244 gaggcctgattcgc 1303
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 961 catactgcttaggaatgc 1020
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1304 catactgcttaggaatgc 1363
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1021 actaagatcctcgc 1080
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1364 actaagatcctcgc 1423
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1081 cctgtgactctgc 1140
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1424 cctgtgactctgc 1483
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1141 ccagacacacttcagctggc 1173
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1484 ccagacacacttcagctggc 1516
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 3
 AAD02697
 ID AAD02697 standard; DNA; 160552 BP.
 XX

AC AD02697;
 XX 02-MAY-2001 (first entry)
 XX Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
 DE Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
 XX therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ds.
 XX Homo sapiens.
 OS
 XX
 XX Location/Qualifiers
 FH Key 32847..32922
 FT exon /*tag= a
 FT /number= 1
 FT /label= 4a_504
 FT intron 32923..35592
 FT /*tag= b
 FT /cons_splice= (5'site:NO, 3'site:YES)
 FT 35593..35674
 FT /*tag= c
 FT /number= 2
 FT /label= 4a_503
 FT intron 35675..45093
 FT /*tag= d
 FT 45094..45185
 FT /*tag= e
 FT /number= 3
 FT /label= 4a_502
 FT intron 45186..46633
 FT /*tag= f
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT 46634..46700
 FT /*tag= g
 FT /number= 4
 FT /label= 4a_501
 FT intron 46701..47938
 FT /*tag= h
 FT /cons_splice= (5'site:YES, 3'site:NO)
 FT 47939..47955
 FT /*tag= i
 FT /number= 5
 FT /note= "Includes 17 base pairs of 5'UTR, the ORF
 FT and all of 3'UTR"
 FT 47939..47955
 FT /*tag= j
 FT /note= "Portion of 5' untranslated region (5'UTR)"
 FT 47956..49128
 FT /*tag= k
 FT /product= "Human glycosyl transferase-4alpha
 FT (GST-4alpha)"
 FT 49129..49746
 FT /*tag= l
 FT 83257..83347
 FT /*tag= m
 FT /label= 4a_502
 FT intron 83348..96412
 FT /*tag= n
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT 96413..96484
 FT /*tag= o
 FT /label= 4a_501
 FT exon 96485..98456
 FT /*tag= p
 FT /cons_splice= (5'site:NO, 3'site:NO)
 FT intron

FT exon 98457..99968
 FT /*tag= q
 FT /note= "Includes 17 base pairs of 5'UTR, the ORF
 FT and all of 3'UTR"
 FT 98457..98473
 FT /*tag= r
 FT /note= "Portion of 5' untranslated region (5'UTR)"
 FT 98474..99661
 FT /*tag= s
 FT /product= "Human glycosyl transferase-4beta
 FT (GST-4beta)"
 FT 99662..99968
 FT /*tag= t
 FT WO200106015-A1.
 PN 25-JAN-2001.
 XX 19-JUL-2000; 2000WO-US19741.
 XX 20-JUL-1999; 99US-0144694.
 PR 13-JUL-2000; 2000US-0593828.
 XX (REGC) UNIV CALIFORNIA.
 PA Rosen SD, Lee JK, Hemmerich S;
 XX WPI; 2001-138471/14.
 DR P-PSDB; AAY72639, AAY72640.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications -
 PS Example 1; Page 62-104; 128pp; English.
 XX The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
 CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
 CC chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, distress
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 XX Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;
 SQ
 Query Match 100.0%; Score 1173; DB 22; Length 160552;
 Best Local Similarity 100.0%; Pred. NO. 3.3e-217;
 Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgtgggtgccacggttctccagcaagacagtgacagtgctctctctggtccagaccacc 60
 Db 47956 atgtgggtgccacggttctccagcaagacagtgacagtgctctctggtccagaccacc 48015
 QY 61 tgcctcctgctcttcatcatctcccgccaggccctcatcccccagcgccgagat 120
 Db 48016 tgcctcctgctcttcatcatctcccgccaggccctcatcccccagcgccgagat 48075
 QY 121 cgtgtgcacgtgctggtgctgctctgctgctcagctcggtcctcttcttggccagctc 180
 Db 48076 cgtgtgcacgtgctggtgctgctctgctgctcagctcggtcctcttcttggccagctc 48135

Best Local Similarity	92.7%;	Pred. No. 2.9e-189;	Mismatches	83;	Indels	3;	Gaps	1;
Matches	1090;	Conservative	0;					
Qy	1	atgtggtgcaacggttctccagcaagacagtgacagtgctctctctctgacagaccac	60					
Db	1159	atgtggtgcaacggttctccagcaagacagtgacagtgctctctctgacagaccac	1218					
Qy	61	tgcctcctgcttctcatctcccgccagcgccctctatcccccagccgagcgagat	120					
Db	1219	tgcctcctgcttctcatctcccgccagcgccctctatcccccagccgagcgagat	1278					
Qy	121	cgtgtgcaacggtgctgctctctctgctgctgctgctgctgctgctgctgctg	180					
Db	1279	cgtgtgcaacggtgctgctctctctgctgctgctgctgctgctgctgctgctg	1338					
Qy	181	tccagcagcaccgacgttctctacgtgatggagcccgctggcatgtgtggacaac	240					
Db	1339	tccagcagcaccgacgttctctacgtgatggagcccgctggcatgtgtggacaac	1398					
Qy	241	ctgtcagggcagcgccgcaacgctgcaacatggccgctgagccacgtgatgctctatc	300					
Db	1399	ctgtcagggcagcgccgcaacgctgcaacatggccgctgagccacgtgatgctctatc	1458					
Qy	301	tttttgcacatgacgctgtgttgatgctctacatgccacagag---ccgaaacctgtcc	357					
Db	1459	tttttgcacatgacgctgtgttgatgctctacatgccacagag---ccgaaacctgtcc	1518					
Qy	358	gcctttttcactgagcaagcagcgccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	417					
Db	1519	gcctttttcactgagcaagcagcgccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1578					
Qy	418	ccccgagcaccatcagcaagcagcgatgtgcaagacacgtgtgcaagcagcaccatc	477					
Db	1579	ccccgagcaccatcagcaagcagcgatgtgcaagacacgtgtgcaagcagcaccatc	1638					
Qy	478	agcctgcccggggagcctgctgctctacagccacggtgtgtgtgtgtgtgtgtgtgt	537					
Db	1639	agcctgcccggggagcctgctgctctacagccacggtgtgtgtgtgtgtgtgtgtgt	1698					
Qy	538	ttcaacctgacgtcccttaccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	597					
Db	1699	ttcaacctgacgtcccttaccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1758					
Qy	598	cacctgtgtgcgaccccgccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	657					
Db	1759	cacctgtgtgcgaccccgccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1818					
Qy	658	gcacgcacacggcagcgt	717					
Db	1819	gcacgcacacggcagcgt	1878					
Qy	718	ctcgcctgattcgcaggt	777					
Db	1879	ctcgcctgattcgcaggt	1938					
Qy	778	aagccgcacaccttctgt	837					
Db	1939	aagccgcacaccttctgt	1998					
Qy	838	gagccgctggcagagatccgcacactacgtctacccgcttccacgctgacccacag	897					
Db	1999	gagccgctggcagagatccgcacactacgtctacccgcttccacgctgacccacag	2058					
Qy	898	ctcagagccttgatcccaacatccacacgggtgtgtgtgtgtgtgtgtgtgtgtgtgt	957					
Db	2059	ctcagagccttgatcccaacatccacacgggtgtgtgtgtgtgtgtgtgtgtgtgtgt	2118					
Qy	958	ttccatacttgcttagaagtcgcgaacagcttccagagcctgagccacgctgtgcc	1017					
Db	2119	ttccatacttgcttagaagtcgcgaacagcttccagagcctgagccacgctgtgcc	2178					
Qy	1018	ttcactaagatcctgctgacagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1077					

Db	2179	ttcactaagatcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	2238
Qy	1078	cggcctgtgtactctgcgaccagcagctgacccacccctgacatctgtgtgtgtgtgtgtgt	1137
Db	2239	cggcctgtgtactctgcgaccagcagctgacccacccctgacatctgtgtgtgtgtgtgtgt	2298
Qy	1138	ggcccaagacacttcaactgctggcctgctgctgctgctgctgctgctgctgctgctg	1173
Db	2299	ggcccaagacacttcaactgctggcctgctgctgctgctgctgctgctgctgctgctg	2334
RESULT 5			
AAD02700 standard; cDNA; 1694 BP.			
XX	AAD02700;		
AC	02-MAY-2001 (first entry)		
XX	Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.		
DE	Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;		
XX	therapy; selectin binding inhibitor; gene therapy; inflammation;		
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;		
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;		
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;		
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anacemia;		
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;		
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;		
XX	chromosome 16q23.1; ss.		
OS	Homo sapiens.		
XX	Key		
PH	5'UTR	Location/Qualifiers	
FT		9..188	
FT	CDS	/*tag= a	
FT		189..1376	
FT		/*tag= b	
FT		/product= "Human glycosyl sulfotransferase-4beta	
FT		(GST-4beta)"	
FT		/note= "CDS is referred as SEQ ID NO:21 in brief	
FT		description of the figures (page no: 4)"	
FT	3'UTR	1377..1694	
FT		/*tag= c	
XX	WC200106015-A1.		
XX	25-JAN-2001.		
XX	19-JUL-2000; 2000WO-US19741.		
XX	20-JUL-1999; 99US-0144694.		
PR	13-JUL-2000; 2000US-0593828.		
XX	(REGC) UNIV CALIFORNIA.		
XX	Rosen SD, Lee JK, Hemmerich S;		
PI	WPI: 2001-138471/14.		
XX	P-PSDB; AAY72640.		
DR	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for		
XX	diagnostic and therapeutic agent screening applications		
PS	Claim 6; Fig 4A; 128pp; English.		
CC	The present sequence is human glycosyl sulfotransferase-4beta (GST-4		
CC	beta) cDNA. GST-4 gene is found on chromosome 16q23.1.		
CC	GST is a type 2 membrane protein useful for inhibiting a binding event		
CC	between a selectin and a selectin ligand, which comprises contacting the		
CC	selectin with a non-sulphated selectin ligand, GST and a small molecular		
CC	agent that inhibits the sulphation activity of GST. GST is also useful		

XX	Human and murine glycosyl sulfotransferase 3 and related polynucleotides -
PT	
PT	

cc gastrointestinal disorders, genetic disorders, immunological
cc disorders, neurological disorders, reproductive disorders, and
cc smooth muscle disorders.

XX sequence 2065 BP: 466 A; 575 C; 491 G; 533 T; 0 other;
CO

Query Match	32.1%	Score 377;	DB 21;	Length 20657
Best Local Similarity	63.4%	Pred. No. 3.9e-64;		
Matches 644; Conservative		0; Mismatches 360; Indels		11

[illegible]

Db 1187 gcctatgaaagggtttctgacttcagaaagccctgtggcgatgcatgaattctggg 1246

QY 1074 ctaccggcctgttactctgcggaccagcagcgtgacctcaacctggatctggtg 1128

Db 1247 ctaccgcacgtcagatctgaacagaaacacagagaacctgttgcgtgattctcg 1301

DECEMBER 9

RESOL
AAZ20793
ID AAZ20793 standard: DNA: 1926 BP.

AA
AC
AA720793:XX
DE
08-DEC-1999 (first entry)

XX
XX

XX Glycacyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondarily lymph organ; ss.
KW

XX	Misc	CD
XX		

XX

XX

FD 30 SEP 1955.
YY

26-FEB-1999; 04

PR 20-MAR-1998;

1. **Introduction**
 2. **Methodology**
 3. **Results**
 4. **Discussion**
 5. **Conclusion**
 6. **References**
 7. **Appendix**
 8. **Index**
 9. **Table of Contents**
 10. **Figure 1**
 11. **Figure 2**
 12. **Figure 3**
 13. **Figure 4**
 14. **Figure 5**
 15. **Figure 6**
 16. **Figure 7**
 17. **Figure 8**
 18. **Figure 9**
 19. **Figure 10**
 20. **Figure 11**
 21. **Figure 12**
 22. **Figure 13**
 23. **Figure 14**
 24. **Figure 15**
 25. **Figure 16**
 26. **Figure 17**
 27. **Figure 18**
 28. **Figure 19**
 29. **Figure 20**
 30. **Figure 21**
 31. **Figure 22**
 32. **Figure 23**
 33. **Figure 24**
 34. **Figure 25**
 35. **Figure 26**
 36. **Figure 27**
 37. **Figure 28**
 38. **Figure 29**
 39. **Figure 30**
 40. **Figure 31**
 41. **Figure 32**
 42. **Figure 33**
 43. **Figure 34**
 44. **Figure 35**
 45. **Figure 36**
 46. **Figure 37**
 47. **Figure 38**
 48. **Figure 39**
 49. **Figure 40**
 50. **Figure 41**
 51. **Figure 42**
 52. **Figure 43**
 53. **Figure 44**
 54. **Figure 45**
 55. **Figure 46**
 56. **Figure 47**
 57. **Figure 48**
 58. **Figure 49**
 59. **Figure 50**
 60. **Figure 51**
 61. **Figure 52**
 62. **Figure 53**
 63. **Figure 54**
 64. **Figure 55**
 65. **Figure 56**
 66. **Figure 57**
 67. **Figure 58**
 68. **Figure 59**
 69. **Figure 60**
 70. **Figure 61**
 71. **Figure 62**
 72. **Figure 63**
 73. **Figure 64**
 74. **Figure 65**
 75. **Figure 66**
 76. **Figure 67**
 77. **Figure 68**
 78. **Figure 69**
 79. **Figure 70**
 80. **Figure 71**
 81. **Figure 72**
 82. **Figure 73**
 83. **Figure 74**
 84. **Figure 75**
 85. **Figure 76**
 86. **Figure 77**
 87. **Figure 78**
 88. **Figure 79**
 89. **Figure 80**
 90. **Figure 81**
 91. **Figure 82**
 92. **Figure 83**
 93. **Figure 84**
 94. **Figure 85**
 95. **Figure 86**
 96. **Figure 87**
 97. **Figure 88**
 98. **Figure 89**
 99. **Figure 90**
 100. **Figure 91**
 101. **Figure 92**
 102. **Figure 93**
 103. **Figure 94**
 104. **Figure 95**
 105. **Figure 96**
 106. **Figure 97**
 107. **Figure 98**
 108. **Figure 99**
 109. **Figure 100**
 110. **Figure 101**
 111. **Figure 102**
 112. **Figure 103**
 113. **Figure 104**
 114. **Figure 105**
 115. **Figure 106**
 116. **Figure 107**
 117. **Figure 108**
 118. **Figure 109**
 119. **Figure 110**
 120. **Figure 111**
 121. **Figure 112**
 122. **Figure 113**
 123. **Figure 114**
 124. **Figure 115**
 125. **Figure 116**
 126. **Figure 117**
 127. **Figure 118**
 128. **Figure 119**
 129. **Figure 120**
 130. **Figure 121**
 131. **Figure 122**
 132. **Figure 123**
 133. **Figure 124**
 134. **Figure 125**
 135. **Figure 126**
 136. **Figure 127**
 137. **Figure 128**
 138. **Figure 129**
 139. **Figure 130**
 140. **Figure 131**
 141. **Figure 132**
 142. **Figure 133**
 143. **Figure 134**
 144. **Figure 135**
 145. **Figure 136**
 146. **Figure 137**
 147. **Figure 138**
 148. **Figure 139**
 149. **Figure 140**
 150. **Figure 141**
 151. **Figure 142**
 152. **Figure 143**
 153. **Figure 144**
 154. **Figure 145**
 155. **Figure 146**
 156. **Figure 147**
 157. **Figure 148**
 158. **Figure 149**
 159. **Figure 150**
 160. **Figure 151**
 161. **Figure 152**
 162. **Figure 153**
 163. **Figure 154**
 164. **Figure 155**
 165. **Figure 156**
 166. **Figure 157**
 167. **Figure 158**
 168. **Figure 159**
 169. **Figure 160**
 170. **Figure 161**
 171. **Figure 162**
 172. **Figure 163**
 173. **Figure 164**
 174. **Figure 165**
 175. **Figure 166**
 176. **Figure 167**
 177. **Figure 168**
 178. **Figure 169**
 179. **Figure 170**
 180. **Figure 171**
 181. **Figure 172**
 182. **Figure 173**
 183. **Figure 174**
 184. **Figure 175**
 185. **Figure 176**
 186. **Figure 177**
 187. **Figure 178**
 188. **Figure 179**
 189. **Figure 180**
 190. **Figure 181**
 191. **Figure 182**
 192. **Figure 183**
 193. **Figure 184**
 194. **Figure 185**
 195. **Figure 186**
 196. **Figure 187**
 197. **Figure 188**
 198. **Figure 189**
 199. **Figure 190**
 200. **Figure 191**
 201. **Figure 192**
 202. **Figure 193**
 203. **Figure 194**
 204. **Figure 195**
 205. **Figure 196**
 206. **Figure 197**
 207. **Figure 198**
 208. **Figure 199**
 209. **Figure 200**
 210. **Figure 201**
 211. **Figure 202**
 212. **Figure 203**
 213. **Figure 204**
 214. **Figure 205**
 215. **Figure 206**
 216. **Figure 207**
 217. **Figure 208**

PA (KEGC) ONLY
PA (SYNT) SYNT

XX	1
XX	2
XX	3
XX	4
XX	5
XX	6
XX	7
XX	8
XX	9
XX	10
XX	11
XX	12
XX	13
XX	14
XX	15
XX	16
XX	17
XX	18
XX	19
XX	20
XX	21
XX	22
XX	23
XX	24
XX	25
XX	26
XX	27
XX	28
XX	29
XX	30
XX	31
XX	32
XX	33
XX	34
XX	35
XX	36
XX	37
XX	38
XX	39
XX	40
XX	41
XX	42
XX	43
XX	44
XX	45
XX	46
XX	47
XX	48
XX	49
XX	50
XX	51
XX	52
XX	53
XX	54
XX	55
XX	56
XX	57
XX	58
XX	59
XX	60
XX	61
XX	62
XX	63
XX	64
XX	65
XX	66
XX	67
XX	68
XX	69
XX	70
XX	71
XX	72
XX	73
XX	74
XX	75
XX	76
XX	77
XX	78
XX	79
XX	80
XX	81
XX	82
XX	83
XX	84
XX	85
XX	86
XX	87
XX	88
XX	89
XX	90
XX	91
XX	92
XX	93
XX	94
XX	95
XX	96
XX	97
XX	98
XX	99
XX	100

XX

DR P-PSDB; AAY3

Human and mu

polymer

PS Claim 4; Fig.

CC This sequence

CC
CC
CC
CC

cc The sequence

cc including co

CC to secondary:

Sequence 19

Query Match

Matches 614;

0y 124 qtqcac

rh 528 atccat

100

Figure 1

22

247

Db 648 accago

QY	304	t t g t g c a c a t g g a c g t g t t t g a t c c t a c a t g --- c c a c a g a g c c g a a a c t g t c o c c c	360
D b	708	c t g t g t a c a t g a g c t g t t t g a t c c t a t g a a c c c a g g c c c c g g a a c a g f t c c a g e	767
QY	361	t t t t t c a a c t g g g c a a c a g a g c o g c g t g t g t c g c c g c c c t g c a g g c c t t t c c c	420
D b	768	c t c t c c a g t g g a g c a a g c g g c c c t g t g t c a g c g c t g t g t g a c t t t c c c t	827
QY	421	c g a g c a c c a t c a a c a a g a g a g t a t g a a g a c a c t g t g c a c g g a c c a g c a c c a t t c a g e	480
D b	828	g c c c a c a g a t c a g c t a c c c a a g c a t g c a a g t g t c t g c g g t c a g c a c c t t i g a t	887
QY	481	c t g c c c g g g a g c t g c g c t c t c a g c a c a g t g t g t c a a g a c c c g t c a a c c t g c a t c g t g c a c	540
D b	888	a t g g t g g a a g g c t g c g c t c a c g g t t g t g t a c c a g g a g g t c g t t t c t c	947
QY	541	a a c c t c a a g t g t c t a c c c g t c t c a g a c c a g t g t g t c a a g a c c c g t c a a c c t g c a t c g t g c a c	600
D b	948	a g c c t c a g c c c t c t a c c a c t a c t c a c g a c c t t c c c t c a a c t c a a c t g a c g t g t g c a c	1007
QY	601	c t g t g c g a c c c g g c g t g c t c g c t c c g g a g c g c g g g c c c g a t a c t a c t g g c a	660
D b	1008	c t g t c c g a a c c c c g g c g t g t c g a t c c c g g a c a c c a c c a t a g a a c t a t g	1067
QY	661	c g c a c a c g y a t g t g t g g g c a c c a a c g g c a a g t g g t g a g c c g a c c c a c c t c a c c t g	720
D b	1068	g t r a c a g t a t t g t c t a g g c a c a t t t g a a c a t c a g a g a g a a c c a g c c c	1127
QY	721	c g c t g a t t c g a g t g t c c g a c c a c g t g c g a t c g c g a g c g g a g c c a c a c t c a a g	780
D b	1128	t a t t a t g c a t a a g a t c a t c t g c a a a g c a a g t g g a c a t a g t c a a g c c a t c c a a c c	1187
QY	781	c c g c a c c c t t c t c g c g c c a c c c t a c c c t g t g c g t t c g a g a c c t g c g c g g g a g	840
D b	1188	c t c c t g a a g c t c t g a g a g c g c t c c t g t c t g a t a g a g a c c t g g t t c g g g c a	1247
QY	841	c c g t g c a g a t c c g c a c t c t a c g c t t c a c c g c t c a c c c t a c c c t a c g c a c a g c t c	900
D b	1248	c c c t g g c c a g a c c a g a c t a t a a t t g t g g g t g g a t t t t t g c c c a c c t c	1307
QY	901	g a g c t g t a t c a c a a c a c c a c c a g g t g g g g a t c g g c a a g c c a a c a t a g a g c c t c	960
D b	1308	c a a a c a t g g t t a c a a t g c a c c c g c g c a a g g c a g g t g c a g c - - - - - a t c c t c	1361
QY	961	c a t a c t t g t a g g a a t g c g c a a c t c t c c a a g c t g g c c a a c c a c c a t a g a g c c t c	1020
D b	1362	c a t a c a a g c c a g a a g c c c a c c t c t c a g g t g a g g t a c g g t g c g t g t g t c t a c t t a c	1421
QY	1021	a c t a a g a t c t c g c g t c a g a g t g t g c c c g c g c t c c a a g c t g g c c a c c g t t g c c t c	1080
D b	1422	g a a a g t t t c c a g t t c a a g a t c t c g g t a g g c t a c g a t g c t g a c t g c t g g t a c a c g	1481
QY	1081	c c t g t a c t c t c g a c c a g a g c a g t g a c c t a c c c t g a t c t g g t	1128
D b	1482	c a g t c a g a t c t c a a g a a a a g g c a a c c t g c c c t g g a t t c t g	1529
RESULT	10		
AAV21200			
ID	AAV21200	standard; cDNA; 2156 BP.	
XX	AAV21200;		
AC			
XX			
DT	01-JUL-1998	(first entry)	
DE	Glycosaminoglycan sulphotransferase gene.		
XX			
KW	Glycosaminoglycan sulphotransferase; C6ST; N-acetylgalactosamine; ds;		
KW	galactose; glycosaminoglycan; chondroitin; keratan sulphate; diagnostic.		
OS	Homo sapiens.		
Key	Location/Qualifiers		
HH			

FT	CDS	147..1586	
FT	/**tag=	a	
FT	/product=	"Glucosaminoglycan sulphotransferase"	
XX			
PN	EP821066-A1.		
XX			
PD	28-JAN-1998.		
XX			
PF	22-JUL-1997;	97EP-0305476.	
XX			
PR	24-JUL-1996;	96JP-0195063.	
XX			
PA	(SEGK) SEIKAGAKU KOGYO CO LTD.		
XX			
PI	Fukuta M, Habuchi O;		
XX			
DR	WPI; 1998-102622/10.		
XX			
PT	Human glycosamino:glycan sulpho:transferase protein - catalyses		
PT	6-sulphation of chondroitin to produce chondroitin sulphate for use		
PT	in pharmaceutical(s)		
XX			
PS	Claim 12; Page 18-21; 27pp; English.		
XX			
CC	The human glycosaminoglycan sulphotransferase (C6ST) gene encodes a type		
CC	II membrane protein (AAW52863) with a molecular weight of 50-55 kD. The		
CC	protein transfers sulphate groups from a donor to the		
CC	N-acetylgalactosamine residue or galactose residue of glycosaminoglycan.		
CC	C6ST displays substrate specificity transferring the sulphate group to the		
CC	hydroxyl group position at C-6 of the N-acetylgalactosamine residue		
CC	of chondroitin and the hydroxyl group position at C-6 of the galactose		
CC	residue of keratan sulphate. The protein is useful for studying the		
CC	function of chondroitin sulphate and can provide chondroitin sulphate		
CC	that may be useful in pharmaceuticals. Both the protein and DNA might		
CC	be useful for treating or diagnosing diseases attributable to low C-6		
CC	sulphation of chondroitin N-acetylgalactosamine residues.		
XX			
SQ	Sequence 2156 BP; 409 A; 695 C; 692 G; 370 T; 0 other;		

[illegible]

Db	930	gagcctgcccgcgcgaaggagacatgcccctcaagcgtgcgcatccgcgcagctggag	989
QY	550	gtgctctaccctgctctcagaccocgcgtcaactcgtgcacactcgtgcactcgtgcgc	609
Db	990	ttcttgagcgcgtgcccgaagaccocgcgtgacctgcgctcatccagctgctgcgc	1049
QY	610	gaccgcgggcgtgctgctgtccgcggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	669
Db	1050	gaccocgcgcgtgctgctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1109
QY	670	ggcctgctgctgctggcaccacgcgaagtgggtggagggccgcacccctcagctgcgtgatt	729
Db	1110	aagaagtgcgtgagcagcagggccagcagcgcgtgagggagagaggtgcagcgcgtg	1169
QY	730	cgccaggtgctgc	789
Db	1170	cggggcacactgcgaag--catccgcctgtccgcggagctgggctgcgcgcgcgcgcgc	1229
QY	790	ttctgcgcgcgcgtaccgcctgtgctgcttcgaggaacctgagcgcgcgcgcgcgcgcgc	849
Db	1227	tggctgcggggccgtacatgctgtgctgacgagcagtgagcgcgcgcgcgcgcgcgcgc	1286
QY	850	ggatcgcgcgcctctacccctcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	909
Db	1287	aaggccgcgcgcgtacccgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1346
QY	910	atccacacatcaccacgc	969
Db	1347	atccaaagacacgc	1397
QY	970	tctaggaaTgc	1029
Db	1398	cagaagaactccgc	1457
QY	1030	ctgc	1089
Db	1458	cagtggtgc	1517
QY	1090	tctgc	1149
Db	1518	gacgc	1577
QY	1150	ttcagctggg 1159	
Db	1578	gtcagctagg 1587	
RESULT 11			
AAT45037			
ID	AAT45037 standard; cDNA to mRNA; 2354 BP.		
XX	AC	AAT45037;	
XX	DT	24-FEB-1997 (first entry)	
XX	DE	Chick chondroitin 6-sulphotransferase cDNA.	
XX	KW	Chondroitin 6-sulphotransferase; C6ST; chondroitin sulphate;	
XX	OS	Gallus sp.	
XX	Key	Location/Qualifiers	
FT	CDS	211..1587	
FT		/*tag= a	
FT	sig_peptide	211..309	
FT		/*tag= b	
FT	mat_peptide	310..1584	
FT		/*tag= c	
FT	misc_RNA	211..1584	
FT		/*tag= d	
FT		/*note= "preferred sequence for use in prodn.	

[illegible]

RESULT	12
AAHI17922	
ID	AAHI17922 standard; cDNA; 3029 Bp.
XX	XX
XX	AC
AAHI17922;	
XX	XX
26-JUN-2001	(first entry)
DT	DT
DE	Human cDNA sequence SEQ ID NO:17678.
XX	XX
Human;	primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW	XX
XX	OS
XX	XX
XX	PN
EP1074617-A2.	
XX	XX
07-FEB-2001.	
PD	XX
XX	XX
28-JUL-2000;	2000EP-0116126.
PF	PF

29-JUL-1999; 99JP-0248036.
27-AUG-1999; 99JP-0300253.
11-JAN-2000; 2000JP-0118776.
02-MAY-2000; 2000JP-0183767.
09-JUN-2000; 2000JP-0241899.
(HELI-) HELIX RES INST.
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama I, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.
Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
Claim 8; SEQ ID 17678; 2537pp + CD ROM; English.
The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
AAB95993 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.
Sequence 3029 BP: 543 A: 945 C; 978 G; 563 T; 0 other;

```

Query Match      13.1%; Score 153.8; DB 22; Length 3029;
Best Local Similarity 50.9%; pred. No. 3.6e-21;
Matches 568; Conservative 0; Mismatches 457; Indels 90; Gaps 5;

```

[illegible]

RESULT	13	
AAx87821		
ID	AAx87821	standard; cDNA; 2409 BP.
XX		
XX	AAx87821;	
XX		
XX	09-NOV-1999	(first entry)
DT		
XX		
DE	Human N-acetylglucosamine-6-O-sulfotransferase cDNA.	
XX		
XX	N-acetylglucosamine-6-O-sulfotransferase; human; GlyCAM-1;	
KW	1-selectin ligand; ss.	
KW		

	XX	Key	Location/Qualifiers
FH		CDS	390..1844
FT			/*tag= a
FT			

XX	EP943688-A2.	
PN		
XX	22-SEP-1999.	
PD		
XX	04-MAR-1999; 99EP-0301530.	
PF		
XX	24-JUN-1998; 98JP-0177844.	
PR		
XX	05-MAR-1998; 98JP-0054007.	
PP		
XX	(SEKK) SEIKAGAKU CORP.	
PA		
XX	Habuchi O, Kadomatsu K, Kaunagi R, Muramatsu H;	
PI	Muramatsu T, Uchimura K;	
PI		
XX	WPI; 1999-520337/44.	
DR	P-PSDB; AAY31657.	
DR		
XX		
XX	New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful	
PT	for synthesis of sugar chains, e.g. GLyCAM-1	
PT		
XX	Claim 7; Page 25-28; 41pp; English.	
XX		
XX	This is the nucleotide sequence of an isolated cDNA coding for	
CC	human N-acetylglucosamine-6-O-sulfotransferase (see AAY31656), an	
CC	enzyme capable of transferring a sulfate group from a sulfate group	
CC	donor to a hydroxyl group at the 6 position of an N-acetylglucosamine	
CC	residue located at the non-reducing end of an oligosaccharide	
CC	represented by the formula GlcNAc β Gal β 1-3Gal β 1-4GlcNAc, where	
CC	beta 1-3 = beta 1-3 glycosidic linkage, Gal = galactose residue,	
CC	beta 1-4 = beta 1-4 glycosidic linkage, and beta 1-4 = beta 1-4	
CC	glycosidic linkage. The cDNA was isolated from a foetal brain	
CC	cDNA library by PCR amplification. The enzyme is useful for the	
CC	synthesis of sugar chains such as GlyCAM-1, a ligand of L-selectin	
CC	that is involved in homing of lymphocytes and rolling of leukocytes	
CC	occurring at the early stage of inflammation. DNA encoding the	
CC	enzyme is expected to be used for the large-scale production of	
CC	N-acetylglucosamine-6-O-sulfotransferase, or artificial synthesis	
CC	of GlyCAM-1 using transformants which harbour the DNA.	
XX	Sequence 2409 BP; 472 A; 741 C; 734 G; 462 T; 0 other;	
XX		

Query Match	12.7%;	Score 148.8;	DB 20;	Length 2409;
Best Local Similarity	50.7%;	Pred. NO. 3.3e-20;		
Matches 563: Conservative	0;	Mismatches 457;	Indels 90;	Gaps 5;

QY	106	gocggcgaggatcgtgtgacagtgctgtgctctcgtggcgctcggctcatcc	165
Db	726	ggcgtcgggacaagcgcactgatgtacgtgttcaacagtggcgtctgctcgtcg	785
QY	166	tctctggcgcaagctcttcagcagcaccccgacgtcttctacctgatggagccgcgttg	225
Db	786	tctctcggcgactatcaaccagaatcccgaggtgtcttctctacgagccagtgtgg	845
QY	226	catgtgtgacacacctgtcgcaggcagcgggcaacgtgcacatggccgtgcgcgac	285
Db	846	catgtatgcaaaactgtatccggggagccggttccctgcaggggcagcgcgcggac	905
QY	286	ctgatgcgctctatctttttgcgacatgacgtgttttgatgctcatgcc-----	338
Db	906	atcctagcgcctcttaccgcgtgcgacctctgtctctccagttgtatgcccgcgggc	965
QY	339	-----acagcgcgaacctgtccgccttttccaactgggcaacgacccgcgcgtg	390
Db	966	agcggggggcgaaacctccaccgcgtggcattctcgcgcagccaccaacaaggtgggtg	1025
QY	391	tgtcgcgcgcgcctgcagccctttcccgaggcacccatcagcaagcagcaagtatgc	450
Db	1026	tgtcgtcaacctctgcccccctacgcgaaggaggtcgtggtgtgtgagcaccgc	1085
QY	451	aag---acactgtgcagcggcagccattcagcctggcccggaggcgctgcgctctcatc	507

Habuchi O.	Kadomatsu K.	Kannagi R.	Muramatsu H:
PPI	Muranatsu T.	Ochimura K;	
XX	WPI: 1999-520337/44.		
DR	P-PSDB; AAY31656.		
DR			
XX	New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful		
PT	for synthesis of sugar chains, e.g. GlyCAM-1		
PP			
XX	Claim 5; Page 21-23; 4lpp: English.		
PS			
XX	This is the nucleotide sequence of an isolated cDNA coding for		
XX	mouse N-acetylglucosamine-6-O-sulfotransferase (see AA31656), an		
CC	enzyme capable of transferring a sulfate group from a sulfate group		
CC	donor to a hydroxyl group at the 6 position of an N-acetylglucosamine		
CC	residue located at the non-reducing end of an oligosaccharide		
CC	represented by the formula GlcNAcbeta1-3Galbeta1-4GlcNAc, where		
CC	GlcNAc = N-acetylglucosamine residue, Gal = galactose residue,		
CC	beta 1-3 = beta 1-3 glycosidic linkage, and beta 1-4 = beta 1-4		
CC	glycosidic linkage. The cDNA was isolated from a mouse embryo		
CC	cDNA library by PCR amplification. The enzyme is useful for the		
CC	synthesis of sugar chains such as GlyCAM-1, a ligand of L-selectin		
CC	that is involved in homing of lymphocytes and rolling of leukocytes		
CC	occurring at the early stage of inflammation. DNA encoding the		
CC	enzyme is expected to be used for the large-scale production of		
CC	N-acetylglucosamine-6-O-sulfotransferase, or artificial synthesis		
CC	of GlyCAM-1 using transformants which harbour the DNA.		
XX			
XX	Sequence 2150 BP; 386 A; 695 C; 679 G; 390 T; 0 other:		
SQ			

Query Match	11.38:	Score 133:	DB 20:	Length 2150:
Best Local Similarity	54.38:	Pred. No. 3.5e-17:		
Matches	324:	Conservative	0:	Mismatches 255: Indels 18: Gaps
QY	109	ggcgcgaggatcgtgtgcacgtgcgtgtgtctcctcgtgcgtcggcgtcggcgtcaccctc	168	
Db	806	ggagggaacaggcagctgtgtgtatgtttcacacgtggcgtcggcgtcgtcttc	865	
QY	169	tgggcgcagctttcagccagcaccccacgtcttctacctgatggagccgcgtggcat	228	
Db	866	tctcgtcgtcttcaacagaccctcagggtgtttcttctctatgacctgtglggcac	925	
QY	229	gtgtggacacacctctgcagcggcagcgcgaacgctgcacatggcgtgcgcacctg	288	
Db	926	ggtgtgcaaaaactgacccggggacgcggtttcctcgcagggggcagcgcggacatg	985	
QY	289	atgcgtctctatcttttgcgcacatggacgtgtttgatgcctacatgc-----	338	
Db	986	ctgagcgtctctaacgcgtgcgtattcttcggttttccacctgtatagcccgaggcagt	1045	
QY	339	----acagagccgaacctctccgccttttcaactgggcgaacgcgcgcgtctgc	393	
Db	1046	ggggggcgcaacctcaccactctgggcattcttggggcagccaactacaaggtggtatgc	1105	
QY	394	tgcgcgcgcgcgtgcgcgcctttcccgaggcaccatcagcaagcagcagctatgcaag	453	
Db	1106	tctcgcacctctgctcgtcactaccgaaaggagtgctgcgtgtgtgacgaccgcgtg	1165	
QY	454	---acactgtgcgcgcgcagccattcagctgcgcggagcgtgcgcctcctcctaacgc	510	
Db	1166	tgcataaagtgccacctcaacgccttgccgcttcggaggaggtgctcgaagtaccgc	1225	
QY	511	cacgtggtcctcaggaggtgcctcttcaacctgcagtgctctaacgcgtgcctcagc	570	
Db	1226	acggtggttatcaaggcgtgcggtctcgtatggtgctgttggcgcgcgtgccttaaa	1285	
QY	571	gaaccgcgtcaacctgcacatcgtgcacctgtgcgcacccgcggcgctgctgcgc	630	
Db	1286	gattccagccttgacctcaaggctcaccactagtacgtgatcctgctgtgttgccagc	1345	
QY	631	tcccggaggcggcgcccgatctctggcagcgcgaacgcgcgtcgtggcacc	687	

Fri Feb 1 20:20:56 2002

us-09-593-828-4.rng

Page 17

Db 1346 tccgcacccgctcggtcaccggtccatccgggaaagccctacaggtggtgcgaagc 1402

Search completed: January 31, 2002, 18:19:26
Job time: 9120 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 16:44:48 ; Search time 3220.32 Seconds

(without alignments)

3914.145 Million cell updates/sec

Title: US-09-593-828-4

Perfect score: 1173

Sequence: 1 atgtgctgcacggtctc.....gctggcgatcgctgactga 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pin: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	657.4	56.0	695	11	BE857538
C 2	631	53.8	735	11	BE858652
C 3	525	44.8	525	11	BE197521
C 4	487.6	41.6	620	10	AL1824100
C 5	420.8	35.9	436	10	AW081348
C 6	352.4	30.0	1923	12	AK009113
C 7	346.4	29.5	1067	13	CNS03KN7
C 8	290.6	24.8	568	10	AL115260
C 9	278	23.7	849	13	CNS04QFN
C 10	271.2	23.1	517	11	BF042384
C 11	240.2	20.5	433	13	AZ405100
C 12	224.4	19.1	695	11	BE964671

13	224.4	19.1	783	11	BE963298
14	218.8	18.7	852	11	BE966340
15	159.4	13.6	965	11	BF597946
16	148.6	12.7	322	10	AI156825
17	134.6	11.5	2070	12	AK011202
18	134.4	11.5	793	11	BI102274
19	130.6	11.1	955	11	BE107354
20	129.2	11.0	634	10	AI180328
21	116.2	9.9	662	11	BF344303
22	111.2	9.5	1070	11	BF163765
23	109.2	9.3	540	10	AW412223
24	103	8.8	103	10	AI282873
25	103	8.8	297	10	AA261202
26	103	8.8	500	10	AI529474
27	99.8	8.5	135	10	AI824198
28	97.4	8.3	434	10	AI088880
29	97.4	8.3	539	11	BE593996
30	97.4	8.3	604	11	BE857485
31	97.4	8.3	722	10	AW027448
32	96.6	8.2	714	11	BI115837
33	96.4	8.2	675	10	AI939595
34	94.2	8.0	468	13	AQ939450
35	84.2	7.2	401	13	AQ017697
36	82.2	7.0	607	11	BE971346
37	81.2	6.9	284	10	AA432229
38	78.6	6.7	516	10	BE208539
39	78.6	6.7	965	11	BE794921
40	78.6	6.7	1094	10	AL576297
41	77.2	6.6	503	11	BF323161
42	76	6.5	550	10	BE754360
43	74.6	6.4	542	10	AL120609
44	72.4	6.2	932	13	CNS00720
45	70.8	6.0	581	11	BI346952

ALIGNMENTS

RESULT 1
BE857538/c
LOCUS
DEFINITION
BE857538 695 bp mRNA EST 29-SEP-2000
7901a08.x1 NCI-CGAP-Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.
ACCESSION BE857538 GI:10371664
VERSION EST
KEYWORDS
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 695)
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
TITLE (CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsbbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LUNL, send email to:
info@image.lunl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 431.
Location/Qualifiers
1..695
/organism="Homo sapiens"

FEATURES
source

```
LOCUS       BE858652       735 bp      mRNA          EST          29-SEP-2000
DEFINITION  7901a09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305176 3'
            similar to TR:Q9UBY3 Q9UBY3 N-ACEYLGLUCOSAMINE
            6-O-SULFOTRANSFERASE. ;, mRNA sequence.
ACCESSION   BE858652
VERSION     BE858652.1 GI:10373890
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 735)
            NCI/MINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGP), Tumor Gene Index
            Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            seq primer: -400P from gibco
            High quality sequence stop: 425.
FEATURES             1..735
             Location/Qualifiers
             1..735
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:3305176"
             /clone_lib="NCI_CGAP_Brn23"
             /tissue_type="glioblastoma (pooled)"
             /lab_host="DH10B"
             /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
             modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
             strand cDNA was primed with a Not I - oligo(dT) primer [5',
             TGTTACCAATCTGAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTTTT
             T 3']; double-stranded cDNA was ligated to Eco RI
             adaptors (Pharmacia), digested with Not I and cloned into
             the Not I and Eco RI sites of the modified pT73 vector.
             Library is normalized, and was constructed by Bento
             Soares and M. Fatima Bonaldo."
BASE COUNT      135 a      221 c      253 g      120 t      6 others
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2
```


QY	544	ctgcagg-tgctctaccgcgtctccagcaccgcccgactcaacctggcgatcgtagcact	602
Dd	256	CTGCAGGTTCGCTCTACCCGCTGCTCAGCGACC CGCGCTCAACTGGCATCGTGCACCT	197
QY	603	gggtcgcgaacccgcggccctgtctggcgtccccgggagcgcgcccgcgatctcggcacg	662
Dd	196	GgTgcGCGACCCCGGGCCGCTGCTGCGTCCCGGAGCGCGGCCGATCTGGGCAAG	137
QY	663	cgacaacggcatcgctcgggcaccaacgcaagtggggtggagcgacacctcacctgcg	722
Dd	136	CGACAACGGCATCGTCTGGGCAACCAACGAAGTGSGTGGAGCGCGACCTCACCTGCG	77
QY	723	cctgattcggaggtgtcgcgcagccacgtgcgcatacgcgcagcgccacacctaagcc	782
Dd	76	CCTGATTGCGAGGTGTGCGCGACGCCAAGTGCATCGCGGAGCGCGCACATCAAGCC	17
QY	783	gccacacttccgcgc	798
Dd	16	GCACACCTTCCTGGC	1
RESULT	6		
LOCUS	AK009113	1923 bp mRNA HTC 05-JUL-2001	
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003G18, full insert sequence.		
ACCESSION	AK009113		
VERSION	AK009113.1 GI:12843701		
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;		
AUTHORS	1 (bases 1 to 1923)		
TITLE	Carninci,P. and Hayashizaki,Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Methods in enzymology. 303, 19-44 (1995)		
PUBMED	99279253		
REFERENCE	10349636		
AUTHORS	2 (bases 1 to 1923)		
TITLE	Carninci,P., Shibata,K., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
MEDLINE	Genome research. 10 (10), 1617-1630 (2000)		
PUBMED	20499374		
REFERENCE	3 (bases 1 to 1923)		
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,I., Kashiwagi,K., Fujisake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,K., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)		
MEDLINE	Genome research. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	4 (bases 1 to 1923)		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE	5 (bases 1 to 1923)		
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,H., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoaka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,		

Db 339 ACCAGACACAGCAGCTGGAAGCTGCACATGCTGTGCGGATCTTCTGCTTCGCTTC 398
Qy 304 ttgtcgacatggacgtgtttgatccatcatg-----ccacgagcgcaaacctgtccgcc 360
Db 399 CTGTGTGACATGACGCTTTTGTATGCTTACATGAACCCAGCGCCCGGAAACAGTCCAC 458
Qy 361 ttttcaactgggcaacgagcgcgcgtgtgtctgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db 459 CTCCTTCAGTGGAGCAAAAGCGGCGCTGTGCTCAGCGCCCTGTGTGTGACTTCTCCCT 518
Qy 421 cagggcacatcagcaagcagcagcgtatcaagacactgtgcacgcgcgcgcgcgcgcgc 480
Db 519 GCCCAGCAGATGAGCTGCACCAAGCACTGCAAGCTGCTTGGCGTCAGCAGCCCTTGAT 578
Qy 481 ctggcccgaggcctgcgcctctacacccacgtgtgtctcaagagggtgcctcttc 540
Db 579 AUGGTGGAAGCCCTGCCCTCTCAGGCTTGTGTGTACTAAGGAGGTGCGTTTCTC 638
Qy 541 aacctgcaggtgtctaccgcgtctcagcgacccgcgcgcgcgcgcgcgcgcgcgcgcgc 600
Db 639 AGCTGCAGGCCCTCTATCTACCTACTACTACGACCCCTTCCCTCAACCTGACGCTCGTGCAC 698
Qy 601 ctggtgc 660
Db 699 CTGGTCGGAGACCCCGCGCGGTTCCTGATCCCGGAGCAGACACCATAGAACTCAIG 758
Qy 721 cgcctgattgcgaggtgtgc 780
Db 819 TATTATGCCATGAAGATCATCTGCAAAAGCCAGGTGACATAGTCAAGGCCATCCAAACC 878
Qy 781 cgcgcacctctcgc 840
Db 879 CTCCTTGAAGCTCTGCAGCAGCGTACTCTGCTGCTGAGTATGAGGACCTGTTGGGCA 938
Qy 841 cgcgtgcgagatccgcgcacttaaccccttaaccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 939 CCGCTGGCCAGACGACGACGACTATAATAATTTGTGGGGTTGGAATTTTGTGCCCCACCTC 998
Qy 901 gaggcctgataccaaataccacacgggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Db 999 CAAACATGGGTTTACAATGTCAACCGCGGCAAGGCGATGGGTACG-----ATGCTTTC 1052
Qy 961 catactgtctaggaatgc 1020
Db 1053 CATATACCGCAGGACGCCCTCAACGCTCTCTCAGGCGTGGCGTGGTTCCTTACCTTAC 1112
Qy 1021 actaagatacctgcgcgtcaggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
Db 1113 GAAAGGTTTCCAGCTTCAAGATGCTCCGCTGAGGCTATGGATTTGTGGGATACCTTC 1172
Qy 1081 cctgtgactcgc 1128
Db 1173 CAGGTCAGATCTCAACAGAACAGGACACCTGTCCCTGGATCTCTG 1220

RESULT 7
CNS03KN7/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
033J20 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL248380.1 GI:7969392
GSS; genome survey sequence.
KEYWORDS
Tetraodon nigroviridis.
Tetraodon nigroviridis.
Fukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.
1 (bases 1 to 1067)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1067)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1067)
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1..1067
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="033J20"
/clone_lib="G"
/note="Genoscope sequence ID : CORG033DE10LP1-end : T7"
BASE COUNT 241 a 306 c 323 g 197 t
ORIGIN
Query Match 29.5%; Score 346.4; DB 13; Length 1067;
Best Local Similarity 62.1%; Pred. No. 3e-55;
Matches 564; Conservative 0; Mismatches 341; Indels 3; Gaps 1;
Qy 17 tctccagcaagacagtgaacgtgctcctcctgagcagaccaccgctcctcgtcttca 76
Db 906 TCGGCACCATGATTTTGTGTGTGACCTTCCAGGGAGCCACGCTGATGCTGAGGGGT 847
Qy 77 tcatctccggccagggccctcatcccgagcgggcgaggtatcgtgcacgtgctg 136
Db 846 GGTACTTCCAGTTCAGCCCTCTGCTCTCCCTCCCTCCGCGGGAAGTTCAGTTCCT 787
Qy 137 tctgtcctcgtggcctcgtggtcctccttcttggcgagctcttcagccagcaccccg 196
Db 786 TCGTGTGCTCCCGGCTCGGGCTCGTCTCTCTGGGTGAGGTTCAGTTCAGTTCAGT 727
Qy 197 acgttcttacctgatggagcccggtggcatgtatggaccaccctgtgcagggcgagcg 256
Db 726 CGGTCTTCTATCTATGATGGAGCCCGGGTGGCACGTGTGGAGCCAGCTCCCAAGTGAACG 667
Qy 257 cggcaacgtgcacatggcgtgcgcaacctgatgcgtctctatcttttgcgacatgg 316
Db 666 TTGGGATGCTCCCGATGCTGTGTCAGGATATATTCGAGCATATTCAGTTCGACTTCT 607
Qy 317 acgtgtttgatgcctacatgccacagagccgaacctgtccgcctttttcaactggcaa 376
Db 606 CAGGAATGGAGGCGCTTCTGCTGAAATCACACAGTATCGGAGCTTTCATGTGGAGCC 547
Qy 377 cggagcgcggtgtgtcctccgcgcctgcagcgcttttcccgaggcaccatcagca 436
Db 546 ACAGCAGGGCGCTGTGTGCGCCCGCGCTCTCTCTCACCCCGCGGAGGATGAGCG 487
Qy 437 agcaggcgtatgcaagacactgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 496
Db 486 ATCAGCCCGAGTGTGTTGAAATGCGGTCTCCAAAGGCTGACAGGGGCTAAGAACCGAT 427
Qy 497 gccctctacagcgaactgtgtcgaaggaggtgccttcttcaacctgcaggtgctct 556
Db 426 GCGACACCTACAGCCACGCTGCTGAAATCGTGCCTTCTTCAGCTGAGTCCCTCTGT 367

Fri Feb 1 20:21:00 2002

us-09-593-828-4.rst

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 695)

REFERENCE

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
<http://image.llnl.gov>

Plate: L14M1095 row: f column: 19
 High quality sequence stop: 684.

FEATURES

source

1..695
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4986258"
 /clone_lib="NCI_CGAP_Co24"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.6 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 118 a 216 c 195 g 166 t

BASE COUNT
 ORIGIN

Query Match 19.1%; Score 224.4; DB 11; Length 695;

Best Local Similarity 65.6%; Pred. No. 1.7e-32;

Matches 391; Conservative 0; Mismatches 196; Indels 9; Gaps 4;

QY 67 ctgtcttcattcatctccgcggccagggccctccatccacagccggcggg---cgagatcg 122
 Db 99 CTCCTCATCCATATCGCTGGCCACAGACACCTTTCCAGAGGAGGAGTCCAGAGGCC 158
 QY 123 tgtgcacgtgctgtgtctctctgtgcgtcgagctcatctctctgtggcagcttt 182
 Db 159 CCGTCAATGTCGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 217
 QY 183 cagcagcagcccgagctctctacatgatggagccgcgcgtggtgacacccct 242
 Db 218 CGGGCAGCACCAGCGATGCTTCTACCTGATGGAGCGCTGCTGGCATGTGGATGACTTT 277
 QY 243 gtgcagggcagcgcggcaacgtgcacatggcgtgagcagctgatcgctctatct 302
 Db 278 CACCAGCAGCAGCAGCTGGAAGCTGCACATGGCTGTGCGGATCTTCTGCTTCCGCTTT 337
 QY 303 ttgtgcacatgagcaggtgtttatccctacatg---ccacagagccgaacacctgtccgc 359
 Db 338 CCGTGTGATGATGAGCGCTTTTATGCTCTACATGAACCCAGAGCCCGCGGAACAGTCCAG 397
 QY 360 cttttcaactggcagcagcgcgcgtgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgc 419
 Db 398 CCGTCTCCAGTGGAGCAAGCCGGGCGCTGTGTCTCAGCGCCTGTGTGTGACTTCTTCCC 457
 QY 420 ccgaggcaccatcagcaagcagcagcagcagcagcagcagcagcagcagcagcagcag 479
 Db 458 TGCCCCAGGATACAGCTACCCAGACACCTGCAAGCTGCTCTGCGGTGAGAGCCCTTTGA 517
 QY 480 cctggccggaggc 539
 Db 518 TATGGTGGAGAGGCGCTGCTCTACGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 577
 QY 540 caacctgcagctgtctctaccgcgtgctcagcagacc-cgcgtcaacctgcagcagctgc 598
 Db 578 CAGCTCAGAGGCGCTTATCCACTGCTCAGGAGCCCTTCCCTCAACCTGCAGCTCCGTC 637
 QY 599 acctggtgcgcagcccgccggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 654

Db 638 CACCTGCTCGAGAGCCCGGGCGCTGTCGATCCCGGGAGCCACACACCATTA 693

RESULT 13

BG963298

LOCUS

DEFINITION

783 bp mRNA

EST 12-JUN-2001

BG963298.1 GI:14350935

house mouse.

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 783)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/ILNL at:

<http://image.llnl.gov>

Plate: L14M1095 row: p column: 04

High quality sequence stop: 741.

Location/Qualifiers

1..783

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4982643"

/clone_lib="NCI_CGAP_Co24"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.6 kb. Constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

141 a 236 c 226 g 180 t

BASE COUNT

ORIGIN

Query Match 19.1%; Score 224.4; DB 11; Length 783;

Best Local Similarity 70.7%; Pred. No. 1.7e-32;

Matches 369; Conservative 0; Mismatches 146; Indels 7; Gaps 5;

QY 124 gtgcacgtgctgtgtctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 183
 Db 190 GTGCATGTCTGT 249
 QY 184 agcagcaccgc 243
 Db 250 GGGCAGCACCAGGATGT 309
 QY 244 tcgagggcagc 303
 Db 310 ACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 369
 QY 304 ttgtgcacatggacgt 360
 Db 370 CTGTGTGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 429
 QY 361 ttttcaactgggcaacagc 420
 Db 430 CTCCTCCAGTGGAGCAAGCCGGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 489
 QY 421 ctaggcaccatcagcaagcagcagcagcagcagcagcagcagcagcagcagcagcag 480

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 16:44:36 ; Search time 3220.32 Seconds
(without alignments)
6820.556 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 gctcgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hcc:*
10: gb_estl:*
11: gb_est2:*
12: gb_hcc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
C 1	657.4	32.2	695	11	BE857538		BE857538 7q01a08.x
C 2	631	30.9	735	11	BE858652		BE858652 7q01a09.x
C 3	525	23.7	525	11	BF137521		BF137521 7q04a08.x
C 4	487.6	23.9	620	10	AI824100		AI824100 wj46c01.x
C 5	420.8	20.6	436	10	AW081348		AW081348 xc41b06.x
C 6	352.4	17.2	1923	12	AK009113		AK009113 Mus muscu
C 7	346.4	16.9	1067	13	CNS03KN7		AL248380 Tetraodon
C 8	308.4	15.1	517	11	BF042384		BF042384 Bp250022A
C 9	295.4	14.5	568	10	AI115260		AI115260 ui43c07.y
C 10	279.4	13.7	849	13	CNS04QPN		AL302540 Tetraodon
C 11	259.4	12.7	955	11	BG107354		BG107354 602290740
C 12	245.4	12.0	433	13	AZ405100		AZ405100 1M0173108

C 13	231.4	11.3	337	11	BG960153		BG960153 PM3-CT064
C 14	224.4	11.0	695	11	BG964671		BG964671 602831875
C 15	224.4	11.0	783	11	BG963298		BG963298 602827716
C 16	218.8	10.7	852	11	BG966340		BG966340 602832826
C 17	217.4	10.6	624	13	AQ373399		AQ373399 RPII11-15
C 18	213.8	10.5	560	10	AL589348		AL589348 DXF28451K
C 19	211.4	10.3	452	11	BF725761		BF725761 bx19a11.y
C 20	211.4	10.3	352	10	AW089016		AW089016 dx34d02.x
C 21	210.2	10.3	400	10	AV760391		AV760391 AV760391
C 22	210.2	10.3	388	13	AQ557029		AQ557029 HS_5308_B
C 23	209.2	10.2	412	10	BE062478		BE062478 QV4-BT025
C 24	209.2	10.2	539	13	AQ379787		AQ379787 RPII11-15
C 25	208.4	10.2	1172	12	AF305824		AF305824 Homo sapi
C 26	208.2	10.2	416	10	AW341978		AW341978 hp74d10.x
C 27	207.8	10.2	616	13	AQ554309		AQ554309 RPII11-4
C 28	207.2	10.1	451	10	AV695478		AV695478 AV695478
C 29	207.2	10.1	650	13	AQ530872		AQ530872 RPII11-3
C 30	207.2	10.1	384	10	AL040054		AL040054 DXF28434P
C 31	207.2	10.1	946	11	BG335756		BG335756 60240452
C 32	206.6	10.1	602	13	AQ540860		AQ540860 RPII11-3
C 33	206.4	10.1	345	11	BG222875		BG222875 naf60e11.
C 34	206.4	10.1	544	11	BG739841		BG739841 602630570
C 35	206.4	10.1	544	13	AQ394650		AQ394650 CTDB1-E1-
C 36	205.6	10.1	440	11	BG189911		BG189911 RST8655_A
C 37	205.4	10.0	400	10	AV760389		AV760389 AV760389
C 38	205.4	10.0	550	10	AL138431		AL138431 DXF28434F
C 39	205.4	10.0	599	10	AW970932		AW970932 EST383015
C 40	204.8	10.0	695	11	BG430795		BG430795 602498565
C 41	204.6	10.0	672	13	AQ631465		AQ631465 RPII11-4
C 42	203.8	10.0	327	10	AA483606		AA483606 ne75d02.s
C 43	203.8	10.0	415	13	AQ150411		AQ150411 HS_3202.B
C 44	203.8	10.0	481	10	AW964231		AW964231 EST376304
C 45	203.6	10.0	495	10	AW835445		AW835445 QV0-LT001

ALIGNMENTS

RESULT 1
BE857538/c
LOCUS
DEFINITION
BE857538 695 bp mRNA EST 29-SEP-2000
7q01a08.x1 NCI-CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. i, mRNA sequence.
ACCESSION
BE857538
VERSION
BE857538.1 GI:10371664
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 695)
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
TITLE
(CGAP/BFCAP), Tumor Gene Index
JOURNAL
Unpublished (1998)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@email.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 431.
Location/Qualifiers
1..695
/organism="Homo sapiens"

BASE COUNT	127	a	210	c	247	g	111	t	ORIGIN
Query Match	32.2%	Score	657.4	DB	11	Length	695		
Best Local Similarity	97.6%	Pred. No.	2.6e-66						
Matches	678	Conservative	0	Mismatches	16	Indels	1	Gaps	1
QY	284	ctgctcttcacatctcccgccagggcgccctcattcccacgcgcggcgagagatgctg	343						
Db	695	TTGTACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTTCTTTT	636						
QY	344	cacdtgctgtgtctctctgctggcgtcggttcattctcttcttgggccactctctcgc	403						
Db	635	CACGTGCTGATGCTCTCTCGTGGCGCTCGGGCTCATCC-TCTTGGGACAGCTCTTCAGC	577						
QY	404	cagcaaccgcgcgtcttctactgatgagccgcgctggcctgtgtggaccacctgtcg	463						
Db	576	CAGCACACCGACGCTCTTCTACTGTATGAGAGCCCGCTGGCATCTGTGGACCATCTGTCG	517						
QY	464	caggcagcgcgcgcgcgcgtcacatggcgcgtgcgcgcctgatgcgtctatctcttttg	523						
Db	516	CAGGCGACGGCGGCAACGCTGCATGCGCGTGGCGAGCTGATGGCGCTATCTCTTTTG	457						
QY	524	tgcgacatggacgtgtgttgatgcctacatgcacagagacgcgaacctgtccgccttttc	583						
Db	456	TGCGACATACAGTGTCTGTATGCCTACATGGCACAGAGCGGAACCTGTCCGCGCTATTC	397						
QY	584	aactgggcaacgagccgcgcgtgtgctgcgcgcgcgcctgcagcgccttcccgcaggc	643						
Db	396	AAC TGCGCAACGAGCCGCGCGCTGTGCTGCCTCCGCCCGCTGCAGCGCCTTCCCGCAGGC	337						
QY	644	accatcagaacgagcagctatgcaagacactgtgcacgcgcgcgcacatcagcctgacc	703						
Db	336	ACCATCAGCAAGCAGGACGATATCAAGACACTGTGCACGGCGGACGACATTCAGCCTGCC	277						
QY	704	cgggagcctgcgcctcctcactcagccactgtgtctcaagaggtgcgccttctcaacctg	763						
Db	276	CGGAGAGCCTTCCGCTCTACGCCAGCTGGTGTCTAAGCAGAGGTGCGCTTCTTCAACCTG	217						
QY	764	cagtgctctaacccgtgctcagcgacccgcgcctcaacctgcgcctgcgcctgcgcctggtg	823						
Db	216	CAGTGCTCTTACCCGCTGCTCAGGACCCCGCGCTCAACCTGCGCATCTGTCACCTGCGTG	157						
QY	824	cgcgacccgcgcgcgtgtcgtctccgggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	883						
Db	156	CGCGACCCGCGGCGCTGCTCGCTCCGGAGGCGGCGGCCGCGATACTTGGCAGCGCAC	97						
QY	884	aacgcgcatcgtctgggcacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	943						
Db	96	AACGGCATCGTGGCTGGGCGACCAACGGCAAGTGGGTGGAGCGCGGACCCCTACCTGGCGCTG	37						
QY	944	attcgcgaggtgtccgc	978						

RESULT 2
BE858652/c

LOCUS BE858652 735 bp mRNA
 DEFINITION 7g01a09.x1 NCI-CGAP.Br23 Homo sapiens cDNA clone IMAGE:3305176 3' similar to TC:Q9UB3 Q9UB3 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 ACCESSION BE858652
 VERSION BE858652.1 GI:10373890
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 735)
 NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index
 Unpublished (1998)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: crapbs-r@mail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -400P from Gibco
 Neg. quality sequence stop 425.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

1. 735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3305176"
/clone_lib="NCI_CGAP_Brun
/tissue_type="glioblasto
/lab_host="PH108"
/notes="Organ; Veno; Vec
modified polylinker; Site
strand cDNA was primed w
TGTTCACATCTGAGTGGGACG
3'; double-stranded
adaptors (Pharmacia), c
the Net I and Eco RI site
Library is normalized, an
Scars and Mating Boud

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

DB 11: Length 735;

Query Match	Score 33.1; 25
Best Local Similarity	93.5%; Pred. NO. 2.6e-63;
Matches 677; Conservative	0; Mismatches 45; Indels 2; Gaps 2;

[illegible]

Db	141	GACACGGGATCTGTGTGGSCACCAACGGCACGTGGTGAGGCGCACCCCGGCTCGGC	82
Qy	941	ctgattcggaggtgtgcccagaccacgtgcgcatcccgagggccgcacactcaagccg	1000
Db	81	GTGGTGGCGAGTGTGCCGTAGCCACGTACGCATCGACAGGCGCGCCACACTCAAGCCG	22
Qy	1001	ccacccttctcgcgggcgcg	1021
Db	21	CCACCCCTTTNTGGCGGCGGC	1

RESULT	5
AW081348/c	
LOCUS	
DEFINITION	AW081348 436 bp mRNA EST 14-OCT-1999 XC41b06.x1 NCI-CGAP-Co20 Homo sapiens cDNA clone IMAGE:2586803 3' similar to TR:O88199 O88199 CHONDROITIN 6-SULFOTRANSFERASE.; mRNA sequence.
ACCESSION	AW081348
VERSION	AW081348.1 GI:6036500
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE	Unpublished (1997)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgaps-remal.nih.gov
COMMENT	Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

Tissue Procurement: Christopher A. Moskaluk, M.D.; Ph.D.; MICHAEL R. Emmert-Buck, M.D.; Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center	
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html	
Seq primer: -40UP from Gibco	
High quality sequence stop: 431.	
FEATURES	Location/Qualifiers
source	1..436
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:2586803"
	/clone_lib="NCI-CCAP_Co20"
	/tissue_type="moderately differentiated adenocarcinoma"
	/lab_host="DH10B"
	/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Normalized to Cot 500. Average insert size 1.1kb. Constructed by Life Technologies."
BASE COUNT	64 a 136 c 169 g 57 t
ORIGIN	
Query Match	20.6%; Score 420.8; DB 10; Length 436;
Best Local Similarity	99.3%; Pred. No. 2.6e-39;
Matches 433; Conservative	0; Mismatches 2; Indels 1; Gaps 1;
Qy 581	ttcaactgggcaacagcgcgctgtgctgcgcgcgcgcctgcagcgctttcccca 640
Db 436	TTCAACTGGGCAACGAGCGCGCTGTGCTGCGCGCGCTGAGCGCCCTTCCCGA 377
Qy 641	ggcaaccatcagcaagcaggacgtatgcaagacactgtgcacgcgcagcattcagcctg 700
Db 376	GSCACCATCAGCAAGCAGGACGTATGCAASAGACTGTGCACGCGGAGCATTCAGCCTG 317
Qy 701	gcccgggagcctgcgcgtctccacagccagtggtgctcaaggaggtgcgcttcttcaac 760
Db 316	GCCCGGGAGCCTGCCGCTCCATACAGCCAGTGGTGCTCAAGAGAGTGGCCTTCCTCCAC 257

QY	761	ctgcagg-tgctctaccgcgtgctcagcagcccgccgctcaacctgcacatcgctgcacct	819
Db	256	CTCAGAGTGTGCTCTACCCGCTGCTCAGGACCCCGCGCTCAACCTGGGCACTCGTGCACT	197
QY	820	ggtgcgagaccgcgggcggctgtgcgctcccgaggagcgcgccgatactggcagc	879
Db	196	GGTGCAGGACCCCGCGCGCTGTGTGCTCCCGGAGCGCGCGCGCGGATACCTGGCAGC	137
QY	880	cgacaacgagcatgctgtggtggcaccacaacgcgaagtgggtggagcgccacccctcactgcg	939
Db	136	CGACAACGCGATCGTGTCTGGGCAACCAACGCAAGTGGGTGGAGCGCGACCTCACCCTGCG	77
QY	940	cctgattcgcgaggtgtgccgcagccacgtgcacatccgcgagggcccccacatcaagcc	999
Db	76	CCTGATTCGGAGGTGTGGCGACGCCACGTGGCGATCGCGAGCGCGCCACACTCAAGCC	17
QY	1000	gacacccctctctgcgc	1015
Db	16	GCACCCCTTCTGCGC	1
RESULT	6		
LOCUS	AK009113	1923 bp mRNA H7C	05-JUL-2001
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003G18, full insert sequence.		
ACCESSION	AK009113		
VERSION	AK009113.1	GI:12843701	
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1923)		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Methods in enzymology. 303, 19-44 (1999)		
PUBMED	92779253		
REFERENCE	2 (bases 1 to 1923)		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome research. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3 (bases 1 to 1923)		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4 (bases 1 to 1923)		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE	5 (bases 1 to 1923)		
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,		

Fri Feb 1 20:20:54 2002

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@wustl.wustl.edu
This clone is available royalty-free through LILNL; contact the
IMAGE Consortium (info@image.lilnl.gov) for further information.
MGI:969488
Seq primer: custom primer used
High quality sequence stop: 517.
Location/Qualifiers
1..568
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone_lib="IMAGE:1885164"
/clone_lib="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site:1: DraIII (CACTGCTG);
Site:2: DraIII (CACCATG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TGTGGCCATGCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGCTG, 3' site
CACCATG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAGCTGCG and 3' end primer
CGACTGCGCTCGAGCACA."

BASE COUNT 80 a 189 c 173 g 126 t
ORIGIN
Query Match 14.5%; Score 295.4; DB 10; Length 568;
Best Local Similarity 75.4%; Pred. No. 3.9e-25;
Matches 381; Conservative 0; Mismatches 121; Indels 3; Gaps 1;
QY 202 gccccagcgccgcatgtggtgcacggttctccagcaagacagtgacagtctct 261
DB 67 GGGCCCTAGCAGAGTATGGGGTACCCCGTTTCTCCAGCAGTGTATGCTTCGCTCT 126
QY 262 cctggcagacacacgtctctgtctctctctctctctccgagcagggccctcatccc 321
DB 127 GATGGTACAG---ACTGGCATCTCTGTTCTCTGGTCTCCGGCAAGTGCATCGTCC 183
QY 322 agcggcgagagatgtgtgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 381
DB 184 AGCAGGCTTTGGGGAGCGTGTGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 243
QY 382 ctcttgggcagctctctccagcagcagcagcagcagcagcagcagcagcagcagcag 441
DB 244 CTTCCTGGCCAGCTCTTCAGCCAAACACCCCGATGCTTCTTACCTGATGGAGCCGCTG 303
QY 442 gcatgtgtgac 501
DB 304 GCACGCTGGGATACGTTGTGCGAGGCGAGTCCCGCCGACATCCACATGCGCGTGA 363
QY 502 cctgatgctctatctttttgtgacatgagcagcagcagcagcagcagcagcagcagcag 561
DB 364 CTTGATGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 423
QY 562 ccgaacacgtccgctttttcaactgggcaacgagcagcagcagcagcagcagcagcagcag 621
DB 424 CCGACACATCTCGGATCTCTCCAGTGGGCGGTGAGCGCGGATCTGCTACCTCCGCT 483
QY 622 ctgcagcgccttctccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 681

DB 484 CTGCGAAGCCTTCGCTCGTGGCAACATCAGCAGGAGGTGCTAAGCCTGTGCGC 543
QY 682 gggcgagcattcagcctgagcctggccgg 706
DB 544 AAGCGCGCCTTCGCGCTGGCTCAG 568
RESULT 10
CNS04QFN 849 bp DNA GSS 24-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 129006 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL302540.1 GI:8181872
VERSION GSS: genome survey sequence.
KEYWORDS Tetraodon nigroviridis.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
REFERENCE 1 (bases 1 to 849)
AUTHORS Rost-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, P., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 849)
Rost-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Barnot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, P.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 849)
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1..849
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/clone_lib="G"
/note="Genoscope sequence ID : COBG129BH03SP1-end :
PUC-ori"
BASE COUNT 137 a 248 c 273 g 184 t 7 others
ORIGIN
Query Match 13.7%; Score 279.4; DB 13; Length 849;
Best Local Similarity 60.8%; Pred. No. 2e-23;
Matches 483; Conservative 3; Mismatches 305; Indels 3; Gaps 2;
QY 602 gcgctgtctgc 661
DB 824 GGGCTGTGCTGCCCGCCGCTGCTCTTCACCCCCCGGAGGATGAGCATGAGCATGCC 765
QY 662 gtatcaagacactgtgc 721
DB 764 GAGTCTTGAAATAATGCTGCGCAAGGCTGCGACGGGCTAGAACATGCGACACC 705
QY 722 tacagccactgtgtgtcctgaagaggtgcctctctctcaacctcaggtgtctaccgcgtg 781
DB 704 TACAGCCACGTGTGTGAATCCGTCGCTCTTCTTCTGAGTGGAGTCCCTGTATCCGCTC 645

QY	782	ctcagcgacccgcgctcaacctgcgaatcgtaactgggtgcgcga--cccgcgggcgtt	840
Db	644	CTGCAGGACCCCAACCATTGAACATCGGTTCACCTTGGTCCGGGACCCCTCGGGCGGT	585
QY	841	gctgcgtctccggaggcggcggccccgatatactggcacgcgcacaacggcatcgtgctggg	900
Db	584	RGMAAAWCMAGAGAAGCAGTCCGCCNNAAGCGCTTCGTGCGC--CGACGACGCCGTAFTTT	527
QY	901	caccaacgcgaagtgggtggaggcgcacacctcaactgcgctgatttcgcgagtgtagccg	960
Db	526	GGAGAACAGCGGCATGCGCGCCGCGAGGCGCATCGTAGAGCCCCCGCTTTCTCAAAGGCGC	467
QY	961	cagccactgctgcgcatcgccgaggcgcaccacacaaacccgccacctctcgtgcgcgcgcg	1020
Db	466	CAGCCACGTGCGCATCAGCAGAGAGGCGCACGCTGAAGGCCCGCCCTTTCTCAAAGGCGC	407
QY	1021	ctaccgcttgtgcgctttcaggacctggcgcggggagccgctggcagagatccgcgaact	1080
Db	406	CTACAAAATGGTCCGCTTCGAGGACCTGGTGCGGACCCCGCTCGGGAGATCGCGGCGCAT	347
QY	1081	ctacgctctcaacgcgctgaacctcagccacacagctcgaggcctggatccacacacatcac	1140
Db	346	CTACAAGTTCGTCGGTCTGGAGATGACCAAGCAGCTGGGCGACGTGATTTACCAAGTAAC	287
QY	1141	ccacggtcggggatcggccaagccaatcgaggcttccatactctgtatggaatgcgcg	1200
Db	286	CCACGGCAAAAGGCACGGGCACCAAGAAGACGCGCTTCGCCATCACCTCCAGAGCGCGC	227
QY	1201	caacgtctccaggcctggcgcacgcgcttgcccttctaagaatcgtgcgcgtcgagga	1260
Db	226	CGACGTCCTCCCAGGCTTGGCGTTCGCGGCTGCGCCACCAAGGTCAAAGCATCCAGGA	167
QY	1261	ggtgtgcgcgcgcgcgtcagctgactgggtfaacgcgcgtgtgtactctgggaccagca	1320
Db	166	ACTGTGCGGGGGCCATCAGCGCTTTGGATACAGGACCGCTGGCAGTGA AAAAAGACA	107
QY	1321	gcgtgacctaaccttgatctggtgtgcacagggccacaccacttcagctgggcatc	1380
Db	106	GAAAGAAGCTAAGCTCGACCTGCTGGTGCCCCAGGAACCATACAGGTTCTCCTGGTTACC	47
QY	1381	gcctgactgagaac	1394
Db	46	GTCCAAGACGGAAC	33
RESULT	11	BGI07354	EST
LOCUS	BGI07354	602290740F1 NIH_MGC_85	Homo sapiens cDNA clone IMAGE:4385888 5'
DEFINITION	mRNA sequence.		
ACCESSION	BGI07354		
VERSION	BGI07354.1		
KEYWORDS	EST.		
SOURCE	human..		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgc.nhl.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov cDNA Library Preparation: Louis Staudt, M.D., Ph.D. cDNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10667 row: o column: 09 High quality sequence stop: 634. Location/Qualifiers		
FEATURES			

and Wright, D., Weiss R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0173 row: I column: 08
Seq primer: CACACGAAGAACGCTATGACC
Class: plasmid ends
High quality sequence stop: 433.

```

FEATURES
source
1. 433
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="JUGCI.M0173108"
/clone_lib="Mouse 10kb p
/sex="Male"
/lab_host="E. Coli strain
/note="vector: PWD42nv;
musculus C57BL/6J (male)
Laboratory Mouse DNA Res
(http://www.jax.org/reso
was hydrodynamically she
0.005 inch orifice at co
was blunt end-repaired v
polynucleotide kinase. I
ligated to the blunt end
adapted DNA was purifi
10.5 kb range using prefi
electrophoresis. Vector
of PWD42 (gi14732114)gb
inducible derivative of
with adaptors compleme
purified. The sheared, a
adapted vector DNA, a
chemically-competent E.
and selected for ampli
112 a 124 c 137 g

```

	Query Match	12.0%	Score 245.4	DB 13	Length 433
	Best Local Similarity	74.7%	pred. No. 2.2e-19		
	Matches 322	Conservative 0	Mismatches 106	Indels 3	Gaps 1
Qy	200	tggccccagccgcccgcgaatgtgctgccagcggtctccagacaagacagtcacagtgctc	259		
Db	428	TAGGCCCTAGCCAGAGTATCGGCTACCGGTTCTCCAGACNGTCATGCTTTCGCTC	369		
Qy	260	ctcctggcacagaccacctgcctcgtctcttcacatctcccgccagggccctcatccc	319		
Db	368	CTGATGTGTACAG--ACNGGCATCTGGCTTCTCTGGTCTCCGCCAAGTGCAITGTCC	312		
Qy	320	ccagccgcgcgaggaatcgtgcaactgctggctgtctctcctcgtgcctcgggctca	379		
Db	311	CCAGCAGGCCTGGGAGCGGTGCACGTCGTGGTACTGCTCCTGGCGCTCGGCGTCG	252		
Qy	380	tctctctgggcagctcttcagccagcaaccgcagctcttctactaatgagccgcg	439		
Db	251	TCCTTCGTGGCGCAGCTTTTCAGCCAAACCCCGATGTTTTCCTGTGAAGCAGCGGCT	192		
Qy	440	tggcatgtgtgaccacccctgctcagggcagcgggcaacgctgcacatggcctggcg	499		
Db	191	TGGCAGCTCTGGGATACGTTGTGCGAGGGCAGTGGCCCCCGGACTCCACATGGCCGTGGT	132		

QY	500	gacctgatcgctctatctttttgtcgacatggacgtgtttgatgcctacatgccacag	559
Db	131	GACCTGATCGCTCAGTGTTCCTATCGACATGGAGCATTTTGAUGCCCTACCTGCCCTGG	72
QY	560	agccgaacactgtccgcctttttcaactggggcaacgacgcgcgcgtgtgctcgccgc	619
Db	71	CGCGACACATCTCGGATCTCTTCAGTGGCGGTGAGCGCGCATTTGCTCACCTCG	12
QY	620	gcctgcagcgc	630
Db	11	GTCTGCGAAGC	1
RESULT	13		
LOCUS	BG960153	337 bp	EST
DEFINITION	PM3-CT0640-300301-003-cl2 CT0640 Homo sapiens cDNA, mRNA sequence.		12-JUN-2001
ACCESSION	BG960153		
VERSION	BG960153.1	GI:14378324	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 337) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed		
JOURNAL	sequence tags		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
COMMENT	20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM3&t2=PM3-CT0640-1) 300301-003-cl2&t3=2001-03-30&t4=1) Seq primer: puc 18 forward High quality sequence start: 20 High quality sequence stop: 126. Location/Qualifiers 1..337 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT0640" /dev_stage="Adult" /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 199716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
FEATURES	source		
BASE COUNT	96 a	87 g	70 t
ORIGIN			
Query Match	11.3%	Score 231.4;	DB 11; Length 337;
Best Local Similarity	91.0%;	Pred. No. 9.8e-18;	
Matches 292;	Conservative	0;	Mismatches 21; Indels 8; Gaps 4
QY	1502	ggagttgggggtccctcctgaagtaagcaaggactgcacgtttttctctctctgattc	1561


```

Db 337 GGAGTTTGGGGTCTCCCTGAAAGTAGCAAGGACTGCACGTTCTTCTCTCTCTGATC 278
QY 1562 tcggtttcccttgagttcttgagctgccttc-----tcacaggtgcactcttcacg 1616
Db 277 TCGGTTTTCCCTTGGAGTCTTGGAGCTGCGCTTCGTCATGACAGGTGCACCTCTTTCATG 218
QY 1617 gaaagcaactc-ttgccctacactctcttgaggagagagtaagtactgc-taaatt 1674
Db 217 GAAATGCAACTCTTTGCCCCACCTCTCTCTGGGCGCAGGAGTAAGTACTGCTTAAT 158
QY 1675 aaattaaatgtgtgcagggcggttgaggtgctcactgcctgtaataccagcattttgag 1734
Db 157 AAATTAATGTGTGCAGCAGGTTACAGTGCCTCATGCTGTAATCCAGCATTTGAG 98
QY 1735 aggtgcaggggtgatcactgaggtcagga-ttcaaaaccagctgcgcaacatagt 1793
Db 97 AGGCTGAGCGGGTGGATCACTGAGGTTCAGGAGTTCGAACCCAGCTGGCCCAACATAGT 38
QY 1794 gaaacccctctctactaaa 1814
Db 37 GATACCCCTCTCTACTAACA 17

RESULT 14
BG964671 695 bp mRNA EST 12-JUN-2001
LOCUS 502831875F1 NCL_CGAP_Co24 Mus musculus cdna clone IMAGE:4986258 5',
DEFINITION mRNA sequence.
ACCESSION BG964671
VERSION BG964671.1 GI:14352308
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cdna Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10955 row: f column: 19
High quality sequence stop: 684.
FEATURES
Location/Qualifiers
1..695
/organism="Mus musculus"
/db_xref="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4986258"
/lab_host="NCL_CGAP_Co24"
/Note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCL_CGAP Library."
BASE COUNT 118 a 216 c 195 g 166 t
ORIGIN
Query Match 11.0%; Score 224.4; DB 11; Length 695;
Best Local Similarity 65.6%; Pred. No. 4e-17;
Matches 391; Conservative 0; Mismatches 196; Indels 9; Gaps 4;
QY 284 ctgctctctatctctccggccagggccctcatcccccagccggc---cgaggatcg 339
Db 99 CTCTTCATCATATGTCCTGGCCACAGACACCTTTCCACAGAGGAGGAGTCCAGGAGGCC 158

```

```

QY 340 tgtcaagtgctgtgtgtctctctctctctctctctctctctctctctctctctctctct 399
Db 159 CGTGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 217
QY 400 cagccagcaccaccacacgtctctctctctctctctctctctctctctctctctctctct 459
Db 218 CGGCGCAGCACCCGGATGTGTCTTACTTGTGAGGAGCTGCTGCGGATCTCTCTGCTTCCGTCCT 277
QY 460 gtgcagggcgagcgcgcaacgctgcacatggcgctgcgcacatgatgcgactgatgcgctctctct 519
Db 278 CACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 337
QY 520 ttgtgcacatggaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 576
Db 338 CCGTGTGCATGAGCGCTTTGATGCTTACATGACCCAGGCCGCCGCCGCCGCCGCCGCC 397
QY 577 ctttttcaactgggcaacgagcgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 636
Db 398 CCGTGTGCATGAGCGCTTTGATGCTTACATGACCCAGGCCGCCGCCGCCGCCGCCGCC 457
QY 637 ccgagggcaccatcagcaagcagcgatgcacacacactgtgcacgcgcgagccattcag 696
Db 458 TCCCCACGAGTACAGTACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 517
QY 697 cctggccgggagggcctgcctctctctctctctctctctctctctctctctctctctctct 756
Db 518 TAIGTGTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 577
QY 757 caactgcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 815
Db 578 CAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 637
QY 816 acctggtgcgagcgagcgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 871
Db 638 CACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 693

RESULT 15
BG963298 783 bp mRNA EST 12-JUN-2001
LOCUS 602827716F1 NCL_CGAP_Co24 Mus musculus cdna clone IMAGE:4982643 5',
DEFINITION mRNA sequence.
ACCESSION BG963298
VERSION BG963298.1 GI:14350935
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cdna Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10985 row: p column: 04
High quality sequence stop: 741.
FEATURES
Location/Qualifiers
1..783
/organism="Mus musculus"
/db_xref="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4982643"
/lab_host="NCL_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"

```

Fri Feb 1 20:20:54 2002

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 kb. Constructed by Life
 Technologies Note: this is a NCI CGAP Library."

BASE COUNT	141 a	236 c	226 g	180 t
ORIGIN				

Query Match	11.0%;	Score 224.4;	DB 11;	Length 783;
Best Local Similarity	70.7%;	Pred. No. 3.8e-17;		
Matches	369;	Mismatches 146;	Indels 7;	Gaps 5;
QY	341	gtgcaagtgtcgtgtcgtgtcctgtgagcgtcgagcctcatctcttctgggccaagctcttc	400	
DB	190	gtgcattgtctgtgtgtgtcttccctggcggtcaggaacctctcttggggacacgcttttc	249	
QY	401	agccagcaccgcgaactctttaccctgatgagccgcgtggcattgtgtggaccacccgtg	460	
DB	250	ggcagcaccggatgtgttacctgatgatgagcctgcctggcatgtgtgatgactttc	309	
QY	461	tgcgagggcagcgcggcaacgctgcacatggccgtgcgagacctgatgccttatcttt	520	
DB	310	accagcagcagcagcctggaaagctgcacatgctgttcgggaatcttctgcgttcgtcttc	369	
QY	521	ttgtgcacatggacgtgttttgatgcctaatg---ccacagagccgaacacctgtccgcc	577	
DB	370	ctgtgtgcacatgagcgtctttgatgtccctacatgaaccacccagcggccggaaacagctccagc	429	
QY	578	tttttaacgtggcgaacgagcgcgcgctgtgtctgcgcgcgcgcgcgcgcgcgcgcgcgcgc	637	
DB	430	ctctttccagttgggagcaaaagcggccctgtgtctcagcgccctgtgtgtgacttcttccct	489	
QY	638	cgaggcaccatcgaacagcaggaactatgaagacacatgtcacgcgcgcgcgcgcgcgcgcgc	697	
DB	490	gccacagagatcagctcaccac	549	
QY	698	ctgcccgcggagcctgcgcctctcacagcacaagtgtgtctcaa-ggaggtgcg-cttt	755	
DB	550	atgtgtggagagagccttgcgcctctcacggcttcgtgggtactcaacggaggtgcgtattccc	609	
QY	756	tcaacctgcagtgctctaccgcgtgtcgaagcccccgcgcctcaacctg-cgcctcgtg	814	
DB	610	ttagccttcagcgcctcttatccactgtctcagagaccttccctcaacctgccacgctcgig	669	
QY	815	caactgtgtgcgcgacccgcgcgcgtgtgcctccggag	856	
DB	670	caactgtgtccgagacc-cggcgccgtgttccgatcccgag	710	

Search completed: January 31, 2002, 16:44:48
Job time: 5772 sec